Deeper Understanding, Faster Calculation
--Exam P Insights & Shortcuts

9th Edition

by Yufeng Guo

For SOA Exam P/CAS Exam 1

Exam Dates

August 21 - 24, 2007

This manual was last updated May 18, 2007

Table of Contents

Chapter 1 Exam-taking and study strategy

Top Horse, Middle Horse, Weak Horse

Sun Bin was a military strategy genius in ancient China who wrote the book Art of War. He advised General Tian who served the Prince of Qi.

As a hobby, General Tian raced and bet on horses with the Prince of Qi. The race consisted of three rounds. Whoever won two or more rounds won the bet. Whoever won least of the three races lost the race and his money bet on the race.

For many years, General Tian could not win the horse race. General Tian always raced his best horse against Prince Qi's best horse in the first round. Because Prince Qi had more money and could buy the finest horses in the country, his best horse was always faster than General Tian's best horse. As a result, General Tian would lose the first round.

Then in the second round, General Tian raced his second best horse against Prince Qi's second best horse. Once again, Prince Qi's middle horse was better than General Tian's middle horse. General Tian would lose the second round of racing.

Usually General Tian could only win the third round. Prince Qi generally won the first two rounds and General Tian won the final round. The Prince won the overall race and collected the bet.

After Sun Bin was hired as the chief military strategist for General Tian, he suggested the General should try a different approach. According to Sun Bin's advice, in the next race General Tian pitted his worst horse against Prince Qi's best horse.

Although the General lost the first round, he could now race his best horse against Prince Qi's second best or middle horse. General Tian's best horse easily defeated the Prince's slower middle horse. The General's middle horse swept past the Prince's worst horse in the third round and clinched the overall victory and winner's purse at last for General Tian.

Moral of the story:

1. The goal of combating a formidable enemy is to win the war, not to win each individual battle. The goal in taking Exam P is to pass the exam, not to get every individual problem correct.

2. Avoid an all-out fight with the opponent on the front where you are weak and the opponent is strong. Surrender to preserve your resources when the opponent's strength outweighs your weakness. When you encounter the most difficult (surprise) problems on the exam, make an intelligent guess if you must but don't

linger over them. Move on to those problems (repeatable) for which you have a solution framework and will most likely solve successfully.

Truths about Exam P

1. Few employers will take your resume seriously without knowing that you have at least passed Exam P. If you are considering becoming an actuary, take Exam P as early as you feel reasonably prepared. Exam P is the entrance exam to the actuary profession. If you fail but are still interested in an actuary career, take the exam again. The sooner you pass Exam P, the more competitive you are in your job search. You might even want to pass Course FM to improve your chances of landing an entry level job.

2. In addition to helping you land a job in the actuary profession, taking Exam P can also help you objectively assess whether you are up to the pressure of actuary exams. To move up in the actuary profession from an apprentice to a fellow, a candidate needs to pass a total of eight exams, beginning with Exam P. Preparing for each exam often requires hundreds of lonely study hours. Taking Exam P will give you a taste of the arduous exam preparation process and helps you decide whether you really want to become an actuary.

   Each year, as new people enter the actuary profession, some actuaries are leaving the profession and finding new careers elsewhere. One of the reasons that people leave the actuary profession is the intense and demanding exam experience. Many intelligent and hard-working people become actuaries, only to find that the exam process is unbearable: too much time away from social activities, too stressful to prepare for an exam, and too painful to fail and re-take an exam.

3. Accept the fact that most likely you won't be 100% ready for Exam P. The scope of study is enormous and SOA has the right to test any randomly chosen concepts from the syllabus. To correctly solve every tested problem, you would have to spend your lifetime preparing for the exam. Accept that you won't be able to solve all the tested problems in an exam. And you certainly do NOT have to solve **all** the problems correctly to pass the exam.

4. Your goal should be to pass the exam quickly so you can look for a job, not to score a 10. Most employers look for potential employees who can pass the exam. They don't care whether you got a 6 or 10.

5. Exam P problems consist of two categories – repeatable problems (easy) and surprise problems (tough). Repeatable problems are those tested repeatedly in the past. Examples include finding posterior probabilities, calculating the mean and variance of a random variable, applying exponential distributions, or applying Poisson distribution to a word problem. These problems are the easiest in the exam because you can master them in advance by solving previous Course 1 or P

exams. Surprise problems, on the other hand, are the brand new types of problems unseen in previous Course 1 and P problems. For example, May 2005 P exam has one question about Chebyshev's inequality. This is a brand new problem never tested in the past. For the majority number of exam candidates, this problem was a complete surprise. Unless you happen to study Chebyshev's inequality when you prepare for Exam P, there's no way you can solve it in the heat of the exam. Surprise problems are designed to make the exam as a whole more challenging and interesting.

6. Strive to get 100% correct on repeatable problems. Most problems on Exam P are repeatable problems. If you can solve these routine problems 100% correct, you have a solid foundation for a passing score.

7. When facing a surprise problem, allow yourself one or two minutes to try and figure it out. If you still can't solve it, just smile, guess, and move on. The rule of thumb is that if you can't solve a problem in three minutes, you won't be able to solve it in a reasonable amount of time that will leave you time to finish the repeatable (and easier) problems on the exam. When facing a hopeless problem, never linger. Move on to the next problem.

8. During the exam, regurgitate, do not attempt to invent. You have on average three minutes per problem on the exam. Three minutes is like the blink of an eye in the heat of the exam. In three minutes, most people can, at best, only regurgitate solutions to familiar problems. Even regurgitating solutions to familiar problems can be challenging in the heat of the exam! Most likely, you cannot invent a fresh solution to a previously unseen type of problem. Inventing a solution requires too much thinking and too much time. In fact, if you find yourself having to think too much in the exam, chances are that you may have to take Exam P again.

9. Be a master of regurgitation. Before the exam, solve and resolve Sample P problems and any newly released P exams (if any) till you get them 100% right under the exam like condition. Build a 3-minute solution script for each of the previously tested P problem. This helps you solve all of the repeatable problems in Exam P, setting a solid foundation for passing Exam P.

Why a 3-minute solution script is critical

Creating a script is a common formula for handling many challenging problems. For example, you must speak for fifteen minutes at an important meeting. Realizing that you are not particularly good at making an impromptu speech, you write and rehearse a script as a rough guide to what you want to say in the meeting.

Scripts are effective because they change the daunting task of doing something special on the spur of the moment to a routine task of doing something ordinary.

Mental scripts are needed for Exam P (and higher level exams). Even most of the repeatable problems in an exam require you to have a thorough understanding of the complex, often unintuitive, underlying concepts. Many tested questions also require you to do intensive calculations with a calculator. A single error in the calculation will lead to a wrong answer. The mastery of complex concepts and the ability to do error-free calculations can't take place in three minutes. They need to be learned, practiced, and memorized prior to the exam.

After you have developed a 3-minute mental script for the repeatable problems for Exam P, solving a problem at exam time is simply a reactivation of the preprogrammed conceptual thinking and calculation sequences. Your 3-minute script will enable you to solve a repeatable type of problems in three minutes.

Remember that there's no such thing as outperforming yourself in the exam. You always under-perform and score less than what your knowledge and ability deserve. This is largely due to the tremendous amount of pressure you inevitably feel in the exam. If you don't have a script ready for an exam problem, don't count on solving the problem on the spur of the moment.

How to build a 3-minute solution script:

1. Simplify the solution, from concepts to calculations, into a 3-minute repeatable process. Just as fast food restaurants can deliver hot hamburgers and French fries into customers' hands in a couple of minutes, you'll need to deliver a solution to an exam question in three minutes.

2. Think simple. Just as a fast food restaurant uses picture menus to allow customers to intuitively see menu options and quickly make a choice, you must convert the textbook version of a concept (often highly complex) into a quick and simple equivalent which allows access to the core of a problem. For example, sample space is a complex and unintuitive concept used by pure mathematicians to rigorously define classical probability. If you find sample space a difficult concept, use your own words to define it ("sample space is all possible outcomes").

3. Calculate fast. Conceptually solving a problem is only half the battle. Unless you know how to press your calculator keys to get the final answer, you won't score a point. So for every repeatable problem in the exam, you need to know the precise calculator key sequences for all the calculations involved in that problem. Many people with sufficient knowledge of Exam P concepts will flounder largely because of slow or sloppy calculations.

Why good candidates fail

At each exam sitting, many good candidates who studied hard for Exam P 1 failed. This unsatisfactory result delayed their entry into the actuary profession.

Candidates fail for several major reasons:

1. Memorizing formulas without gaining a sophisticated understanding of the core concepts. Exam P appears formula-driven. It seems that as long as you have memorized the myriad of formulas, you should easily pass the exam. Such erroneous thinking has misled many candidates into an exam disaster. Those who have memorized formulas without understanding the nuances and substance of core concepts walk into the exam room, only to find that subtly designed exam problems render many of their memorized formulas useless.

2. Reading too much, but solving too few practice problems (especially previous SOA Course 1 and Sample P problems). Studying for Exam P is like learning how to swim. While reading a book on how to swim is helpful, ultimately you need to immerse yourself in the water and discover how to swim rather than sink.

   To pass Exam P, you need to do more than read about core concepts. You must immerse yourself in Sample Exam P.

3. Doing busy work instead of learning, progressing, and excelling. If solving too few practice problems can lead to exam failure, the opposite extreme can be true too. A hard-working exam candidate can literally solve hundreds of practice problems (even SOA problems) without learning much and fail the exam miserably.

   Let's see how an untrained swimmer ultimately becomes a world champion. In the beginning the swimmer has many bad swimming habits such as uncoordinated body movements and poor breathing. To become a better athlete he must gradually shed his bad habits and relearn correct swimming techniques from the ground up, perhaps from a good coach. Gradually, he becomes a more graceful, more coordinated, and more efficient swimmer. After perfecting his swimming skills for many years, his technique propels him toward the level of stamina and skill which will enable him to compete for and win a gold medal.

   Now imagine another hard-working swimming who practices twice as much as the first swimmer but who doesn't bother changing his poor swimming habits. He merely swims according to what he knows how to swim. The more he practices swimming, the more entrenched his poor swimming habits become. Years go by and he is still a bad swimmer.

   When doing practice problems for Exam P, focus on shedding your poor habits such as lengthy thinking or error-prone calculation. Focus on gaining conceptual insights and learning efficient computation skills. Learn how to solve problems in a systematic and streamlined fashion. Otherwise, solving hundreds of practice problems only reinforces your bad habits and you'll make the same mistakes over and over.

4. Leisure strong, exam weak. Some candidates are excellent at solving problems leisurely without a time constraint, but are unable to solve a problem in 3 minutes during the heat of an exam. The major weakness is that thinking is too lengthy or calculation is too slow. If you are such a candidate, practice previous SOA exams under a strict time constraint to learn how to cut to the chase.

5. Concept strong, calculation weak. Some candidates are good at diagnosing problems and identifying formulas. However, when it's time to plug the data into a calculator, they make too many mistakes. If you are such a candidate, refer to Chapter Two of this book to learn how to do calculations 100% right 100% of the time.

Recommended study method

1. Sense before you study. For any SOA or CAS exams (Exam P or above), always carefully scrutinize one or two of the latest exams to get a feel for the exam style before opening any textbooks. This prevents you from wasting time trying to master the wrong thing.

   For example, if you look at any previous Sample P problems, you'll find that SOA never asked candidates to define a sample space, a set, or a Poisson distribution. Trying to memorize the precise, rigorous definition of any concept is a complete waste of time for Exam P. SOA is more interested in testing your understanding of a concept than in testing your memorization of the definition of a concept.

2. Quickly go over some textbooks and study the fundamental (the core concepts and formulas). Don't attempt to master the complex problems in the textbooks. Solve some basic problems to enhance your understanding of the core concepts.

3. Focus on applications of theories, not on pure theories. One key difference between the probability problems in Exam P and the probability problems tested in a college statistics class is that SOA problems are oriented toward solving actual problems in the context of measuring and managing insurance risks, while a college exam in a statistics class is often oriented toward pure theory. When you learn a new concept in Exam P readings, always ask yourself, "What's the use of such a concept in insurance? How can actuaries use this concept to measure or manage risks?"

   For example, if you took a statistics class in college, you most likely learned some theories on normal distribution. What you probably didn't learn was why normal distribution is useful for insurance. Normal distribution is a common tool for actuaries to model aggregate loss. If you have N independent identically distributed losses $x_1, x_2, \dots, x_n$, then the sum of these losses

$$S_n = x_1 + x_2 + \dots + x_n$$

is approximately normally distributed when N is not too small, no matter whether individual loss x is normally distributed or not. Because S_n is approximately normally distributed, actuaries can use a normal table and easily find the probability of S_n exceeding a huge loss such as Pr(S_n >$10,000,000).

4. Focus more on learning the common sense behind a complex theorem or a difficult formula. Focus less on learning how to rigorously prove the theorem or derive the formula. One common pitfall of using college textbooks to prepare for Exam P is that college textbooks are often written in the language of pure mathematical statistics. These textbooks tend to place a paramount emphasis on setting up axioms and then rigorously proving a theorem and deriving a formula. Though scholastically interesting, such a purely theoretical approach is unproductive in preparing for SOA and CAS exams.

5. Master sample Exam P Exam and any newly released Exam P (if any). SOA exam problems are the best practice problems. Work and rework each problem till you have mastered it.

CBT (computer-based testing) and its implications

1. There are more exam sittings in 2006 and beyond. 2006 has four sittings for Exam P. If a candidate fails Exam P in one sitting, he can take another exam several months later.

2. Most likely, SOA won't release any CBT exam P. The old Course 1 exams you can download from the SOA website is all you can get. Unless SOA changes its mind, you won't see any Exam P to be released in the near future.

3. CBT contains a few pilot questions that are not graded. You don't know which problem is a pilot and which is not. Pilot questions won't add anything to your grade, even if you have solved them 100% right.

4. When taking CBT, learn to tolerate imperfections of CBT and have a peaceful frame of mind.

    - Your assigned CBT center may be several hours away from your home. Be sure to check your CBT center out several days before the exam.

- Expect to have computer monitor freezing problems. Many candidates reported that their computer monitors froze up, from time to time, for a few seconds.

- Learn to cope with a dim and tiny table at your CBT center. You may be assigned to such a table at your CBT center.

- Learn to tolerate noise in the exam room.

5. Check out the official CBT demo from the SOA website. Get comfortable with CBT. Lear how to navigate from one problem to the next, how to mark and unmark a problem.

Chapter 2 Doing calculations 100% correct 100% of the time

What calculators to use for Exam P

SOA/CAS approved calculators:

BA-35, BA II Plus, BA II Plus Professional, TI-30X, TI-30Xa, TI-30X II (IIS solar or IIB battery).

Best calculators for Exam P: BA II Plus, BA II Plus Professional, TI-30X IIS.

I recommend you buy two calculators: (1) TI-30X IIS (because it's solar-powered, you don't need to worry about it running out of battery), (2) either BA II Plus or BA II Plus Professional.

TI-30X IIS costs about $15. BA II Plus costs about $30. BA II Plus Professional costs from $48 to $70.

If you already have BA II Plus, you don't need the more expensive BA II Plus Professional for Exam P. However, you will need to buy BA II Plus Professional for Exam FM.

BA II Plus Professional is Texas Instrument's new calculator selling from $48 - $70 in retail stores. SOA just added BA II Plus Professional as one of the approved calculators for the 2005 exams.

BA II Plus Professional can do everything BA II Plus does. In addition, BA II Plus has new features currently lacking in BA II Plus. One new feature in BA II Plus Professional is the modified duration calculation. Because the modified duration is on Exam FM syllabus, you'll need a BA II Plus Professional for Exam FM.

You can buy BA II Plus Professional at amazon.com for less than $50 (typically free shipping). This price includes a 1-year subscription to the Money magazine. If you don't want the subscription, you can return a postage-paid card to Money and ask for a refund. The refund is about $14. So your net cost of BA II Plus Professional is about $36.

Be aware that when entering numbers into a BA II Plus Professional calculator, you have to press the keys a lot harder than you do when entering numbers on a BA II Plus or TI-30X IIS. In this respect I find the BA II Plus and TI-30X IIS easier to use than the BA II Plus Professional.
When you buy the BA II Plus (or BA II Plus Professional) and the TI-30X IIS, one tip is to choose one color for the BA II Plus (or Plus Professional) and a strikingly different color for the TI-30X IIS. For example, buy a black BA II Plus, and a purple TI-30X IIS. This way, in the heat of the exam, you know exactly which calculator is which.

I recommend that you spend several hours running the calculation examples listed in the calculator manual (Texas Instruments called it “Guidebook”) for BA II Plus, BA II Plus Professional, and TI-30X IIS. If your calculator didn’t have a guidebook when you bought it, you can go to www.ti.com to download a guidebook.

If you are a borderline student, calculation skills make or break your exam. If you can do messy calculations 100% correct 100% of the time, every problem solved is a point earned. You might have a better chance of passing the exam than someone else who knows more than you about the subject but who makes mistakes here and there.

I recommend that you devote some of your precious study time toward mastering BA II Plus and TI-30X IIS. They are to you in the exam as a weapon is to a soldier in a combat. Many candidates make the mistake of spending too much time learning concepts and too little time learning calculation skills.

The guidebooks for BA II Plus (or Professional) and TI-30X IIS are straightforward and easy to learn, so I don’t want to repeat the guidebooks. I just want to highlight some of the most important things you need to know about BA II Plus (or Professional) and TI-30X IIS.

Because BA II Plus and BA II Plus Professional are equally good for Exam P, I’ll only talk about BA II Plus. Please keep in mind that if BA II Plus can do a job, then BA II Plus Professional can do the same job with identical key strokes. So in the following discussion, if you see the term BA II Plus, it refers to both BA II Plus and BA II Plus Professional.

Contrasting the BA II Plus with TI-30X IIS

Each calculator has its strengths and weaknesses. The first strength of the BA II Plus is that it is faster than TI-30X IIS. Immediately after you enter the data into BA II Plus, it gives you the answer right away. There is no waiting time. In contrast, the TI-30X II takes several seconds before it gives you the final answer, especially when you have a complex formula or have many numbers to insert into a formula.

Second, the BA II Plus often requires fewer key strokes than TI-30X IIS for the same calculation. For example, to calculate $\sqrt{3}=1.73205081$, you have two key strokes for the BA II Plus: “3 $\sqrt{x}$ ” (which means first you press 3 and then press $\sqrt{x}$). However, the TI-30X IIS requires four key strokes: “2^{nd} $\sqrt{\ }$ 3 =” (you first press 2^{nd}, then press $\sqrt{\ }$, then enter “3”, and finally press “=”).

The third strength of BA II Plus is that it has separators but TI-30X IIS does not. For example, if you enter one million into BA II Plus, you’ll see “1,000,000” (there are two separators “,” in this figure to indicate the unit of 1000). However, if you enter one million into TI-30X IIS, you’ll see “1000000.” Lack of separators in TI-30X IIS increases your chances of entering a wrong number into the calculator or reading a wrong output from the calculator if the number entered or displayed is big.

The fourth and greatest advantage of BA II Plus is that it has more powerful statistics functions than the TI-30X IIS. In the one-variable statistics mode, you can enter any positive integer as the data frequency for BA II Plus. In contrast, the maximum data frequency in TI-30 IIS is 99. If the data frequency exceeds 99, you cannot use TI-30X IIS to find the mean and variance. This severely limits the use of the statistics function for the TI-30X IIS.

Please note that the TI-30X IIS can accommodate up to 42 distinct values of a discrete random variable X. BA II Plus Statistics Worksheet can accommodate up to 50 distinct values.

The restriction of no more than 42 data pairs in TI-30X IIS and no more than 50 in BA II Plus is typically not a major concern for us; the exam questions asking for the mean and variance of a discrete random variable most likely have fewer than a dozen distinct values of X and $f(x)$ so candidates can solve each problem in three minutes.

The fifth strength of BA II Plus is that it has 10 memories labeled M0, M1, ...,M8, M9. In contrast, TI-30 IIS has only 5 memories labeled A,B,C,D, and E.

Right now, you may wonder if the TI-30X IIS has any value given its many disadvantages. The power of the TI-30X IIS lies in its ability to display the data and formula entered by the user. This "what you type is what you see feature" allows you to double check the accuracy of your data entry and of your formula. It also allows you to redo calculations with modified data or modified formulas.

For example, if you want to calculate $2e^{-2.5}-1$, as you enter the data in the calculator, you will see the display:

$$2e\wedge(-2.5)\ -1$$

If you want to find out the final result, press the "Enter" key and you will see:

$$2e\wedge(-2.5)\ -1$$
$$-0.835830003$$

So $2e\wedge(-2.5)\ -1=-0.835830003$

After getting the result of -0.835830003, you realize that you made an error in your data entry. Instead of calculating $2e\wedge(-2.5)\ -1$, you really wanted to calculate $2e\wedge(-3.5)\ -1$. To correct the data entry error, you simply change "–2.5" to "–3.5" on your TI-30X IIS. Now you should see:

$$2e\wedge(-3.5)\ -1$$
$$-0.939605233$$

With the online display feature, you can also reuse formulas. For example, a problem requires you to calculate $y = 2e^{-x} - 1$ for $x_1 = 5$, $x_2 = 6$, and $x_3 = 7$. There is no need to do three separate calculations from scratch. You enter $2e^{\wedge}(5) - 1$ into the calculator to calculate y when x=5. Then you modify the formula to calculate y when $x = 6$ and $x = 7$.

This redo calculation feature of TI-30X IIS is extremely useful for solving stock option pricing problems in Course 6 and Course 8V (8V means Course 8 for investment). Such problems require a candidate to use similar formulas to calculate a series of present values of a stock, with the calculated present value as the input for calculating the next present value. Such calculations are too time-consuming using the BA II Plus under exam conditions when you have about 3 minutes to solve a problem. However, these calculations are a simple matter for the TI-30X IIS.

Please note that BA II Plus can let you see the data you entered into the calculator, but only when you are using BA II Plus's built-in worksheets such as TVM (time value of money), Statistics Worksheet, $\Delta\%$ Worksheet, and all other worksheets. However, the BA II Plus does not display data if you are NOT using its built-in worksheets. Unlike the TI-30X IIS, BA II Plus never displays any formulas, whether you are using its built-in worksheets or not.

So which calculator should I use for Exam P, BA II Plus or TI-30X IIS?

Different people have different preferences. You might want to use both calculators when doing practice problems to find out which calculator is better for you.

These are my suggestions:

Statistics calculations --- If you need to calculate the mean and variance of a discrete random variable, use (1) BA II Plus Statistics Worksheet, or (2) TI-30X IIS for its ability to allow you to redo calculations with different formulas. When using TI-30X IIS, you can take advantage of the fact that the formula for the mean and the formula for the variance are very similar:

$$E(X) = \sum x f(x), \quad E(X^2) = \sum x^2 f(x), \quad Var(X) = E(X^2) - E^2(X)$$

So you first calculate $E(X) = \sum x f(x)$, then modify the mean formula to

$E(X^2) = \sum x^2 f(x)$, and calculate the variance $Var(X) = E(X^2) - E^2(X)$.

Don't use the TI-30X IIS statistics function to calculate the mean and variance because the TI-30X IIS statistics function is inferior to the BA II Plus Statistics Worksheet. Later in this book, I will give some examples on how to calculate the mean and variance of a discrete random variable using both calculators.

Other calculations – Most likely, either BA II Plus or TI-30X IIS can do the calculations for you. BA II Plus lets you quickly calculate the results with fewer calculator key strokes, but it doesn't leave an audit trail for you to double check your data entry and formulas (assuming you are not using the BA II Plus built-in worksheets). If you are not sure about the result given by BA II Plus, you have to reenter data and calculate the result a second time. In contrast, TI-30X IIS lets you see and modify your data and formulas, but you have more calculator key strokes and sometimes you have to wait several seconds for the result. I recommend that you try both calculators and find out which calculator is right for you for which calculation tasks.

Critical calculator tips

You might want to read the guidebook to learn how to use both calculators. I'll just highlight some of the most important points.

BA II Plus – always choose AOS. BA II Plus has two calculation methods: the chain method and the algebraic operating system method. Simply put, under the chain method, BA II Plus calculates numbers in the order that you enter them. For example, if you enter $2+3\times 100$, BA II Plus first calculates 2+3=5. It then calculates $5\times 100=500$ as the final result. Under AOS, the calculator follows the standard rules of algebraic hierarchy in its calculation. Under AOS, if you enter $2+3\times 100$, BA II Plus first calculates $3\times 100=300$. And then it calculates $2+300=302$ as the final result.

AOS is more powerful than the chain method. For example, if you want to find $1+2e^{3}+4\sqrt{5}$, under AOS, you need to enter

$$1 + 2\times\ 3\ 2^{\text{nd}}\, e^{x}\ + 4\times\ 5\sqrt{x}\quad \text{(the result is about 50.1153)}$$

Under the chain method, to find $1+2e^{3}+4\sqrt{5}$, you have to enter:

$$1+(2\times 3\ 2^{\text{nd}}\, e^{x})+(4\times 5\sqrt{x})$$

You can see that AOS is better because the calculation sequence under AOS is the same as the calculation sequence in the formula. In contrast, the calculation sequence in the chain method is cumbersome.

BA II Plus and TI-30X IIS – always set the calculator to display at least four decimal places. This way, you'll be able to see the final result of your calculation to at least four decimal places.

BA II Plus can display eight decimal places; TI-30X IIS can display nine decimal places. I recommend that you set both calculators to display, at minimum, four decimal places.

Please note that the internal calculations done in BA II Plus or TI-30X IIS are NOT affected by the number of decimal places you set the calculator to display. Your choice of decimal places only affects what you can see on the calculator screen.

BA II Plus and TI-30X IIS—be careful about storing and retrieving intermediate values. If a calculation has several steps, you will need to store the intermediate values somewhere – either on scrap paper (which will be the exam paper in the actual exam) or in the calculator's memories. Both methods have pros and cons.

If you like to copy intermediate values onto scrap paper (the simpler approach of the two), writing down all the decimal places will be too time-consuming for some heavy-duty calculations and your calculation may lose some precision. If you like to store intermediate values in the calculator's memory, you do not have to transfer intermediate values back and forth between the scrap paper and your calculator. However, you have the extra burden of keeping track of which memory stores which intermediate values.

Mistakes associated with using calculator's memories are prevalent. One common mistake is thinking you stored an intermediate value in your calculator's memory but you didn't, in which case your intermediate value is lost and you have to recalculate. Another common mistake is to store an intermediate value in one memory (such as Memory M0 in BA II Plus or in Memory A in TI-30 IIS) but to retrieve it from another memory (such as Memory M1 in BA II Plus or Memory B in TI-30X IIS).

Watch out for the distinction between the minus sign and the negative sign. Both BA II Plus and TI-30X IIS have a minus sign "－" and a negative sign " - ." They look similar, but are completely different. If you look closely, the minus sign is longer and the negative sign is shorter. The real difference is that the minus sign is for subtraction (such as 5－3=2) and the negative sign is to indicate that a number is negative (such as e^{-1} =0.36787944). Using the subtraction sign mistakenly for the negative sign will cause a calculation error.

Be careful using % on the BA II Plus. For example, if you type 100 – 5% in BA II Plus, you might expect to get 99.95. However, after pressing the [Enter] key, you get 95 instead. BA II Plus calculates $A \pm X\% = A(1 \pm X\%)$. So if you type 100+5%, you'll get 105. If your intention is to calculate 100 – 0.05, you can enter 100 -1ı5% (i.e. 100 minus 1 times 5%).

Learn how to reset your calculator. According to exam rules, when you use BA II Plus or TI-30X IIS for an SOA or CAS exam, exam proctors on site will need to clear the memories of your BA II Plus or TI-30X IIS. Typically, a proctor will clear your calculator's memories by resetting the calculator to its default setting. This is done by pressing "2nd" "Reset" "Enter" for BA II Plus and by simultaneously pressing "On" and "Clear" for TI-30X IIS and TI-30X IIB. You will need to know how to adjust the settings of BA II Plus and TI-30X IIS to your best advantage for the exam.

In its default settings, BA II Plus, among other things, uses the chain calculation method, displays only two decimal places, and sets its Statistics Worksheet to the LIN (standard linear regression) mode. After the proctor resets your BA II Plus and before the exam begins, you need to change the settings of your BA II Plus to the AOS calculation method, at least a four decimal place display, and the 1-V (one variable) as the mode for its Statistics Worksheet. Refer to the guidebook on how to change the settings of your BA II Plus.

You need change only one item on your TI-30X IIS once the proctor resets it. In its default settings, TI-30X IIS displays two decimal places. Instead set your TI-30X IIS to display at least four decimal places.

When you study for Exam P, practice how to reset your BA II Plus and TI-30X IIS and how to change their default settings to the settings most beneficial for taking the exam.

Finally … practice, practice, practice until you are perfectly comfortable with the BA II Plus and TI-30X IIS. Work through all the examples in the guidebook relevant to Exam P.

Calculator Exercise 1

A group of 23 highly-talented actuary students in a large insurance company are taking SOA Exam FM at the next exam sitting. The probability for each candidate to pass Course 2 is 0.73, independent of other students passing or failing the exam. The company promises to give each actuary student who passes Course 2 a raise of $2,500. What's the probability that the insurance company will spend at least $50,000 on raises associated with passing Exam FM?

Solution

If the company spends at least $50,000 on exam-related raises, then the number of students who will pass FM must be at least 50,000/2,500=20. So we need to find the probability of having at least 20 students pass FM.

Let X = the number of students who will pass FM. The problem does not specify the distribution of X. So possibly X has a binomial distribution. Let's check the conditions for a binominal distribution:

- There are only two outcomes for each student taking the exam – either Pass or Fail.
- The probability of Pass (0.73) or Not Pass (0.27) remains constant from one student to another.
- The exam result of one student does not affect that of another student.

X satisfies the requirements of a binomial random variable with parameters n =23 and p =0.73. We also need to find the probability of $x \geq 20$.

$$\Pr(x \geq 20) = \Pr(x = 20) + \Pr(x = 21) + \Pr(x = 22) + \Pr(x = 23)$$

Applying the formula $f_X(x) = C_n^x p^x (1-p)^{n-x}$, we have

$$f(x \geq 20)$$
$$= C_{23}^{20}(.73)^{20}(.27)^3 + C_{23}^{21}(0.73)^{21}(.27)^2 + C_{23}^{22}(.73)^{22}(.27) + C_{23}^{23}(.73)^{23} = .09608$$

Therefore, there is a 9.6% of chance that the company will have to spend at least \$50,000 to pay for exam-related raises.

Calculator key sequence for BA II Plus:

Method #1 – direct calculation without leaving audit trails

Procedure	Keystroke	Display
Set to display 8 decimal places (4 decimal places are sufficient, but assume you want to see more decimals)	$\boxed{2^{nd}}$ $\boxed{\text{Format}}$ 8 $\boxed{\text{Enter}}$	DEC=8.00000000
Set AOS (Algebraic operating system)	$\boxed{\text{2nd}}$ [FORMAT], keep pressing $\boxed{\downarrow}$ multiple times until you see "Chn." Press $\boxed{\text{2nd}}$ [ENTER] (if you see "AOS", your calculator is already in AOS, in which case press [CLR Work])	AOS
Calculate $C_{23}^{20}(.73)^{20}(.27)^3$		
Calculate C_{23}^{20}	23 $\boxed{2^{nd}}$ $\boxed{{}_nC_r}$ 20 $\times$	1,771.000000
Calculate $(.73)^{20}$	.73 $\boxed{y^x}$ 20 $\times$	3.27096399
Calculate $(.27)^3$	.27 $\boxed{y^x}$ 3 +	0.064328238
Calculate $C_{23}^{21}(0.73)^{21}(.27)^2$		
Calculate C_{23}^{21}	23 $\boxed{2^{nd}}$ $\boxed{{}_nC_r}$ 21 $\times$	253.0000000
Calculate $(.73)^{21}$	.73 $\boxed{y^x}$ 21 $\times$	0.34111482

Calculate $(.27)^2$	.27 $\boxed{x^2}$ +	0.08924965
Calculate $C_{23}^{22}(0.73)^{22}(.27)$		
Calculate C_{23}^{22}	23 $\boxed{2^{nd}}$ $\boxed{{}_nC_r}$ 22 ×	23.00000000
Calculate $(.73)^{22}$	.73 $\boxed{y^x}$ 22 ×	0.02263762
Calculate $(.27)$	.27 +	0.09536181
Calculate $C_{23}^{23}(0.73)^{23}$		
Calculate C_{23}^{23}	23 $\boxed{2^{nd}}$ $\boxed{{}_nC_r}$ 23 ×	1.00000000
Calculate $(.73)^{23}$ and get the final result	.73 $\boxed{y^x}$ 23 =	0.09608031

Method #2 – With audit trails

Procedure	Keystroke	Display
Set to display 8 decimal places (4 decimal places are sufficient, but assume you want to see more decimals)	$\boxed{2^{nd}}$ $\boxed{\text{Format}}$ 8 $\boxed{\text{Enter}}$	DEC=8.00000000
Set AOS (Algebraic operating system)	$\boxed{\text{2nd}}$ [FORMAT], keep pressing $\boxed{\Downarrow}$ multiple times until you see "Chn." Press $\boxed{\text{2nd}}$ [ENTER] (if you see "AOS", your calculator is already in AOS, in which case press [CLR Work])	AOS
Clear memories	2^{nd} MEM 2^{nd} CLR Work	M0=0.00000000
Get back to calculation mode	CE/C	0.00000000
Calculate $C_{23}^{20}(.73)^{20}(.27)^3$ and store it in Memory 1		
Calculate C_{23}^{20}	23 $\boxed{2^{nd}}$ $\boxed{{}_nC_r}$ 20 ×	1,771.000000
Calculate $(0.73)^{20}$	.73 $\boxed{y^x}$ 20 ×	3.27096399

Calculate $(.27)^3$	.27 $\boxed{y^x}$ 3 =	0.06438238
Store the result in Memory 0	STO 0	0.06438238
Get back to calculation mode	CE/C	0.00000000
Calculate $C_{23}^{21}(0.73)^{21}(.27)^2$ and store it in Memory 1		
Calculate C_{23}^{21}	23 $\boxed{2^{nd}}$ $\boxed{{}_nC_r}$ 21 $\times$	253.0000000
Calculate $(.73)^{21}$	.73 $\boxed{y^x}$ 21 $\times$	0.34111482
Calculate $(.27)^2$	.27 $\boxed{x^2}$ =	0.02486727
Store the result in Memory 1	STO 1	0.02486727
Get back to calculation mode	CE/C	0.00000000
Calculate $C_{23}^{22}(0.73)^{22}(.27)$ and store it in Memory 3		
Calculate C_{23}^{22}	23 $\boxed{2^{nd}}$ $\boxed{{}_nC_r}$ 22 $\times$	23.00000000
Calculate $(.73)^{22}$	.73 $\boxed{y^x}$ 22 $\times$	0.02263762
Calculate $(.27)$	.27 =	0.00611216
Store the result in Memory 2	STO 2	0.00611216
Calculate $C_{23}^{23}(0.73)^{23}$ and store it in Memory 4		
Calculate C_{23}^{23}	23 $\boxed{2^{nd}}$ $\boxed{{}_nC_r}$ 23 $\times$	1.00000000
Calculate $(.73)^{23}$ and get the final result	.73 $\boxed{y^x}$ 23 =	0.00071850
Store the result in Memory 3	STO 3	0.00071850
Recall values stored in Memory 1,2,3, and 4. Sum them up.		
	RCL 0	0.06438238
	+ RCL 1	0.02486727
	+	0.08924965
	RCL 2	0.00611216
	+	0.09536181
	RCL3	0.00071850
	=	0.09608031

Comparing Method #1 with Method #2:

Method #1 is quicker but more risky. Because you don't have an audit history, if you miscalculate one item, you'll need to recalculate everything again from scratch.

Method #2 is slower but leaves a good auditing trail by storing all your intermediate values in your calculator's memories. If you miscalculate one item, you need to recalculate that item alone and reuse the result of other calculations (which are correct).

For example, instead of calculating $C_{23}^{20}(.73)^{20}(.27)^{3}$ as you should, you calculated $C_{23}^{20}(.73)^{3}(.27)^{20}$. To correct this error under method #1, you have to start from scratch and calculate each of the following four items:

$C_{23}^{20}(.73)^{20}(.27)^{3}$, $C_{23}^{21}(0.73)^{21}(.27)^{2}$, $C_{23}^{22}(0.73)^{22}(.27)$, and $C_{23}^{23}(0.73)^{23}$

In contrast, correcting this error under Method #2 is lot easier. You just need to recalculate $C_{23}^{20}(.73)^{20}(.27)^{3}$; you don't need to recalculate any of the following three items:

$$C_{23}^{21}(0.73)^{21}(.27)^{2},\ C_{23}^{22}(0.73)^{22}(.27),\ \text{and}\ C_{23}^{23}(0.73)^{23}$$

You can easily retrieve the above three items from your calculator's memories and calculate the final result:

$$C_{23}^{20}(.73)^{20}(.27)^{3}+C_{23}^{21}(0.73)^{21}(.27)^{2}+C_{23}^{22}(.73)^{22}(.27)+C_{23}^{23}(.73)^{23}=.09608$$

I recommend that you use Method #1 for simple calculations and Method #2 for complex calculations.

<u>Calculator keystrokes for TI-30X IIS:</u>

Now let's solve the same problem using TI-30 IIS. Before entering the formula for $f(x \geq 20)$ into a TI-30X IIS calculator, you might want to calculate the four items:

$$C_{23}^{20}=1{,}771,\ C_{23}^{21}=253,\ C_{23}^{22}=23,\ C_{23}^{23}=1.$$

To calculate the first two items above, you can use the combination operator in BA Plus or TI-30X IIS (BA Plus is faster than TI-30X IIS). For TI-30X IIS, the operator key for C_{23}^{20} is "23 PRB ${}_nC_r$ 20 ENTER." You should be able to calculate the last two items without using a calculator.

Now you are ready to enter the formula:

$$1771(.73\text{^}20)(.27\text{^}3)+253(.73\text{^}21)(.27^2)+23(.73\text{^}22).27+.73\text{^}23$$

Press [ENTER], wait a few seconds, and you should get 0.09608031.

Calculator Exercise 2 -- Fraction math

Evaluate $\frac{1}{2}+\frac{5}{3}+\frac{7}{11}-\frac{2}{9}$. Express your answer in a fraction (i.e. you cannot express it as a decimal).

Solution

Fraction math happens when you do integration in calculus or solve probability problems. It is simple conceptually, but is easy to make a mistake in the heat of the exam.

Here is a shortcut – you can use TI-30X IIS to do fraction math for you. One nice feature of the TI-30X IIS is that it can convert a decimal number into a fraction. This feature can save you a lot of time. For example, if we want to convert 0.015625 into a fraction, TI-30X IIS can quickly do the job for you. The key sequence for TI-30X IIS is:

Type .015625
Press 2nd PRB (PRB key is to the right of LOG key)
Press ENTER

Then TI-30 IIS (IIB) will display 1/64

So 0.015625=1/64
Now, back to our original problem …

To convert $\frac{1}{2}+\frac{5}{3}+\frac{7}{11}-\frac{2}{9}$ into a fraction, the key strokes are:

Type 1/2+5/3+7/11-2/9
Press ENTER
Press 2nd PRB
Press ENTER

TI-30X IIS will give you the result: $2\frac{115}{198}$

Calculator Exercise 3

Evaluate $\int_0^1 \left(\frac{1}{2} + \frac{3}{5}x - x^2 + \frac{11}{6}x^3 - \frac{9}{7}x^4 \right) dx$. Express your result as a fraction.

Solution

$$\int_0^1 \left(\frac{1}{2} + \frac{3}{5}x - x^2 + \frac{11}{6}x^3 - \frac{9}{7}x^4 \right) dx$$

$$= \left[\frac{1}{2}x + \frac{3}{5} \times \frac{1}{2}x^2 - \frac{1}{3}x^3 + \frac{11}{6} \times \frac{1}{4}x^4 - \frac{9}{7} \times \frac{1}{5}x^5 \right]_0^1$$

$$= \frac{1}{2} + \frac{3}{5} \times \frac{1}{2} - \frac{1}{3} + \frac{11}{6} \times \frac{1}{4} - \frac{9}{7} \times \frac{1}{5}$$

For TI-30X IIS, enter

$1/2 + 3/(5*2) - 1/3 + 11/(6*4) - 9/(7*5$

Press ENTER, you should get: 0.667857143
Press 2nd PRB ENTER, you should get: 187/280

So $\int_0^1 (\frac{1}{2} + \frac{3}{5}x - x^2 + \frac{11}{6}x^3 - \frac{9}{7}x^4) dx = \frac{187}{280}$

Please note that for TI-30X IIS to convert a decimal number into a fraction, the denominator of the converted fraction converted must not exceed 1,000. For example, if you press

0.001 2nd PRB ENTER

You will get 1/1000 (so 0.001=1/1000)
However, if you want to convert 0.0011 into a fraction, TI-30X IIS will not convert. If you press:

0.0011 2nd PRB ENTER

You will get 0.0011. TI-30 IIS will not convert 0.0011 into 11/10,000 because 0.0011 =11/10000 (the denominator 10,000 exceeds 1,000). However, this constraint is typically not an issue on the exam.

Comparison of 3 best calculators

Features	**BA II Plus Professional**	**BA II Plus**	**TI-30X IIS**
Approximate cost	\$48-\$70	\$30	\$15
Calculation speed	Fast	Fast	Slower
Have separators?	Yes	Yes	No
Max # of decimals displayed	8	8	9
# of memories	10	10	5
Leave audit trail?	Only in its built-in worksheets	Only in its built-in worksheets	All calculations
1-V Statistics			
# of distinct values	50	50	42
data frequency	any positive integer	Any positive integer	A positive integer not exceeding 99
Ease of entering data	Have to press calculator keys hard	Soft and gentle	Soft and gentle
Good for Exam FM?	Yes. Can calculate Modified Duration. Good for all calculations (including time value of money).	Yes. Cannot calculate Modified Duration, but good for other calculations (including time value of money).	Yes. Not good for time value of money, but good for general calculations
Good for upper exams?	Yes	Yes	Yes
Optimal settings for Exam P	Use AOS; display at least 4 decimals; set statistics to 1-V mode.	Use AOS; display at least 4 decimals; set statistics to 1-V mode.	Display at least 4 decimals

Homework for you – read the BA II Plus guidebook and the TI-30 IIS guidebook. Work through relevant examples in the guidebooks. Rework examples in this chapter.

Chapter 3 Set, sample space, probability models

Probability models use set notation extensively. Let's first have a review of set theories. Then we'll look at how to build probability models.

Set theories

Set. A set is a collection of objects. These objects are called elements or members of the set.

If an object x is an element of the set S, then we say $x \in S$.

Example. The set of all even integers no smaller than 1 and no greater than 9 is $A = \{2, 4, 6, 8\}$. Set A has four elements: 2,4,6, and 8.
$2 \in A; \quad 4 \in A; \quad 6 \in A; \quad 8 \in A.$.

Example. The set of all non-negative even integers is $B = \{0, 2, 4, 6, ...\}$. B has an infinite number of elements.

Null set. A null set or an empty set is a set that has no elements. A null set is often expressed as $\varnothing$. $\varnothing = \{\ \}$.

Subset. For two sets A and B, if every element of A is also an element of B, we say that A is a subset of B. This is expressed as $A \subseteq B$.

Example. Set A contains the even integers no smaller than 1 and no greater than 9. Set B contains all the non-negative even integers. Set C contains all integers. Set D contains all real numbers.

Then $A \subset B; \quad B \subset C; \quad C \subset D.$

Identical set. If every element in set A is an element in set B and every element in set B is also an element in set A, then A, B are two identical sets. This is expressed as $A = B$.
Example. Set A contains all real numbers that satisfy the equation $x^2 - 3x + 2 = 0$. Set $B = \{1, 2\}$. Then $A = B$.

Complement. For two sets A and U where A is a subset of U (i.e. $A \subseteq U$), the set of the elements in U that are not in A is the complement set of A. The complement of A is expressed as $\overline{A}$ (some textbooks express it as A^c).

Example. Let $A=\{1,7\}$ and $U=\{1,2,3,5,7,9\}$. $\overline{A}=\{2,3,5,9\}$.

Union. For two sets A and B, the union of A and B contains all the elements in A and all the elements in B. The union of A and B is expressed as $A\cup B$.

Example. Let $A=\{1,2,3\}$, $B=\{2,5,7,9\}$. $A\cup B=\{1,2,3,5,7,9\}$.

Example. Let set C represent all the even integers no smaller than 1 and no greater than 9; set D represent all non-negative integers. $C\cup D=D$.

Intersection. For two sets A and B, the intersection of A and B contains all the common elements in both A and B. The intersection of A and B is expressed as $A\cap B$.

Example. Let $A=\{1,2,3\}$, $B=\{2,5,7,9\}$. $A\cap B=\{2\}$.

Example. Let $A=\{1,3\}$, $B=\{2,5,7,9\}$. $A\cap B=\varnothing$.

Example. Let set C represent all the even integers no smaller than 1 and no greater than 9, set D represent all non-negative integers. $C\cap D=C$.

Steps to build a probability model

A probability model is a simplification and abstraction of a real world experiment. It gives us a framework to assign probability and analyze a random process.

Step 1 Set up an appropriate sample space
Step 2 Set up basic probability laws (called axioms)
Step 3 Assign probability measures

To understand how to how to build a probability, let's look at some examples.

Case 1 Toss a fair coin

All probability models are concerned about a random process whose outcomes are not unknown in advance. This random process is called an experiment. A probability model attempts to answer the question "What's the chance that we observe one outcome out of many possible outcomes?"

In the case of tossing a coin, we don't know whether heads or tails will show up. So tossing a coin is a random process whose outcomes are not known in advance. We want to find out the chance of getting a head or tail in one toss of coin. We'll build a probability model using the 3-step process.

Step 1 Determine sample space

Sample space is all possible outcomes of an experiment. When setting sample space, we'll want to make sure

- We haven't forgotten any possible outcomes. In other words, we must exhaustively list all of the possible outcomes. (exhaustive property)
- We haven't listed the same outcome twice. In other words, the outcomes we have listed should be mutually exclusive. (mutually exclusive property)

When tossing a coin, we'll get either heads (H) or tails (T). So our sample space is {H,T}. H and T are mutually exclusive yet collectively exhaustive.

Step 2 Set up probability axioms

Before we dive into the details of calculating the probability of certain outcomes, we need to set up and follow some high level probability laws. These laws won't tell us exactly how to calculate a probability. However, by following these laws, we can build a probability model that is consistent and not self-contradictory.

All probability models must follow these 3 axioms (Kolmogorov's axioms):

Axiom 1 Probability is non-negative. For any event A, $P(A) \geq 0$.

Axiom 2 Probability is additive. If A and B are two mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$.

Axiom 3 Probabilities must add up to one. If Ω represent the sample space (which is the list of all possible outcomes), then $P(\Omega)$ =1. In other words, if we perform an experiment, we are 100% certain to observe one of the possible outcomes.

Applying these three axioms to the coin-toss example, we have:

$\Omega = \{H, T\}$

$P(H) \geq 0$, $P(T) \geq 0$

$P(H \cup T) = P(H) + P(T)$ -- because H and T are mutually exclusive

$P(\Omega) = P(H \cup T) = 1$

Please note these axioms don't tell us how to calculate $P(H)$ or $P(T)$. The purpose of the axioms is to make sure that our model is sound. The calculation of $P(H)$ and $P(T)$ requires additional work, in the confines of the three probability axioms.

If you have trouble seeing what probability axioms are good for, think about writing an essay. Your high school teacher told you millions of times that an essay needs to have 3 parts: opening, body, and conclusion. “Having an opening, a body, and a conclusion” is like an axiom for writing essays. This axiom gives you a framework for writing essays. If you follow this axiom, your essay is structurally sound. However, following this axiom won’t guarantee that your essay is good. Your essay can have an opening, a body, and a conclusion, but it may still be a bad essay. To be good, besides having an opening, a body, and a conclusion, your essay needs to have substance.

Similarly, if you follow the three probability axioms, your probability model is structurally sound. However, this doesn’t guarantee that your probability model is good. To be good, your probability model needs to have substance. You must have detailed knowledge of the random process.

Step 3 Assign probability measures

Assigning probability measures requires the detailed knowledge of the experiment. Fortunately, tossing a coin is a simple experiment; our common sense is enough to help us assign probabilities.

It’s reasonable to assume that heads and tails are equally likely to occur for a fair coin.

$$\Rightarrow \quad P(H)=P(T)$$

We combine Step 2 and 3:

$$P(H\cup T)=P(H)+P(T)$$
$$P(\Omega)=P(H\cup T)=1$$
$$P(H)=P(T)$$

Solving the above equations, we have:

$$P(H)=P(T)=\frac{1}{2}$$

General rule:

If an experiment has n outcomes that are equally likely to occur, then the probability of an event A is:

$$P(A)=\frac{\text{\# of elements in } A}{n}$$

Case 2 Toss an unfair coin

In a toss of an unfair coin, Step 1 and Step 2 of the model building process are the same as in a toss of a fair coin. The only difference is on Step 3.

If a coin is not fair, then $P(H) \neq P(T)$. To calculate $P(H)$ and $P(T)$, we need to know more about the coin. For example, if our analysis indicates that heads are twice as likely to occur as tails, then

$$P(H) = 2P(T)$$

We combining Step 2 and 3:

$$P(H \cup T) = P(H) + P(T)$$
$$P(\Omega) = P(H \cup T) = 1$$
$$P(H) = 2P(T)$$

Solving the above equations, we have:

$$P(H) = \frac{2}{3}, \quad P(T) = \frac{1}{3}$$

Step 3 of the model building process requires one to understand the experiment that is being studied. There is no one correct way to assign probability. People with different experiences may assign different probabilities to the same event. However, no matter how you assign your probability, if you follow the 3 axioms of probability, your model is internally consistent.

Case 3 Tossing two fair coins

Step 1 Determine sample space

We face a choice here. Should we consider the two coins to be distinguishable and set up ordered pairs of H and T? Or should we consider them to be indistinguishable and hence set up unordered pairs of H and T? Let's look at each option.

Option 1. We assume the two coins are different. We can think one coin is A and the other is B.

The sample space is:

$$\Omega = \left\{ H^A H^B, H^A T^B, T^A H^B, T^A T^B \right\}$$

$H^A H^B$ represents that Coin A is heads up and B is heads up and so on.
Dropping the superscript, we have:

$$\Omega = \{HH, HT, TH, TT\}$$

Option 2. We assume the two coins are identical. Then "HT" and "TH" are the same because they both give us "one head and one tail." Then we have only three outcomes: two heads, two tails, one head and one tail. The sample space is:

Ω ={ Two heads, Two tails, One head and one tail }

For now let's consider only option 1 (the two coins are distinguishable). The sample space is $\Omega = \{HH, HT, TH, TT\}$. Now let's continue building the model.

Step 2 Apply the 3 axioms

$$P(HH) \geq 0,\ \ P(HT) \geq 0,\ P(TH) \geq 0,\ P(TT) \geq 0$$
$$P(\Omega) = P(HH \cup HT \cup TH \cup TT) = P(HH) + P(HT) + P(TH) + P(TT) = 1$$

Step 3 Assign probability

It's reasonable to assume that each outcome is equally likely to occur. So

$$P(HH) = P(HT) = P(TH) = P(TT)$$

Applying the axiom $P(HH) + P(HT) + P(TH) + P(TT) = 1$, we have:

$$P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$$

Now let's look at the 2nd option of specifying the sample space. Now the two coins are indistinguishable. The sample space is $\Omega = \{HH, HT, TT\}$. Step 2 of the model building process tells us:

$$P(HH) \geq 0,\ \ P(HT) \geq 0,\ P(TT) \geq 0$$
$$P(\Omega) = P(HH \cup HT \cup TT) = P(HH) + P(HT) + P(TT) = 1$$

When we assign probability in Step 3, we have a trouble. Since we treat the two coins are the same, now we can't easily assess the probability of having one heads and one tail. For example, we can't say that it's equally likely to have two heads, two tails, one head and one tail.

This brings up an important point. When specifying the sample space, make sure your outcomes are detailed enough for you to answer the question at hand.

Case 4 Randomly draw a number from [0, 1]

Step 1 Determine sample space

Let x represent the number randomly drawn from the interval [0,1]. Then the sample space is $\Omega = \{x:\ 0 \le x \le 1\}$.

Step 2 Apply axioms

$$P(x) \ge 0, \qquad P(\Omega) = P(0 \le x \le 1) = 1$$

Please note that for the above two equations to hold, we must have $P(x) = 0$ for any $0 \le x \le 1$. If $P(x)$ is a positive number, then when we sum up the total probability from $x = 0$ to $x = 1$, we'll get an infinite number. To have $P(\Omega) = P(0 \le x \le 1) = 1$, we must have $P(x) = 0$.

Step 3 Assign probability

How can we assign probability measure to satisfy the axioms? We can assign probability of having $x \in [a, b]$ (where $0 \le a \le b \le 1$) proportional to the length $b - a$:

$$P(a \le x \le b) = k(b - a) \quad \text{(where } 0 \le a \le b \le 1 \text{ and } k \text{ is a constant)}$$

$$\Rightarrow \quad P(\Omega) = P(0 \le x \le 1) = k(1 - 0) = 1, \quad k = 1$$
$$\Rightarrow \quad P(a \le x \le b) = 1 \times (b - a) = b - a$$
$$\Rightarrow \quad P(x = a) = P(a \le x \le a) = 1 \times (a - a) = 0$$

By assigning $P(a \le x \le b) = b - a$ (where $0 \le a \le b \le 1$), we'll be able to satisfy the three axioms of probability.

Case 5 Randomly draw two numbers x and y from [0, 1]. Find the probability for $x + y \le 0.5$.

Step 1 Determine sample space

$$\Omega = \{(x, y):\ 0 \le x \le 1,\ 0 \le y \le 1\}$$

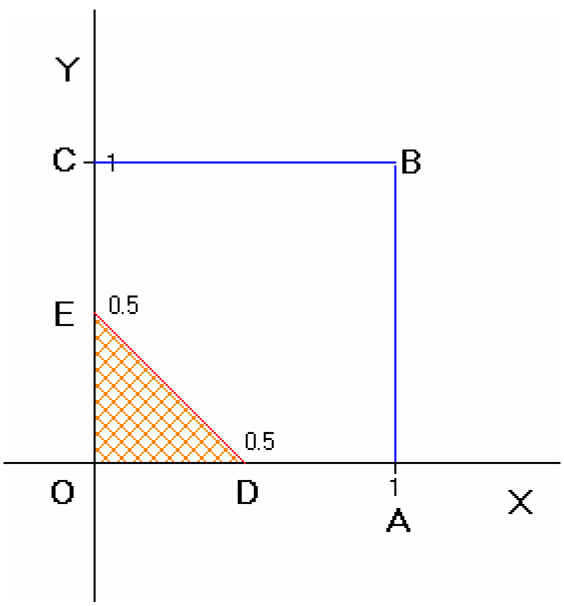
Step 2 Satisfy axioms

$P(x,y) \geq 0$ for $0 \leq x \leq 1,\ 0 \leq y \leq 1$
$P(\Omega) = 1$

Once again, we need to have $P(x,y) = 0$ for any (x,y). If $P(x,y)$ is positive, when we add up the total probability $P(\Omega)$, we'll get an infinite number. So to have $P(\Omega) = 1$, we must have $P(x,y) = 0$ for any (x,y).

Step 3 Assign probability

Here (x,y) lies in the unit square ABCO. To assign probability such that $P(\Omega) = 1$ is satisfied, we can assign $P(0 \leq x + y \leq 1)$ to be proportionally to the area DEO. Line DE represents $x + y = 0.5$; area DEO represents $0 \leq x + y \leq 0.5$.



Since area ABCO corresponds to $P(\Omega) = 1$, we have:

$$P(0 \leq x + y \leq 1) = \frac{\Delta DEO}{\Delta ABCO} P(\Omega) = \frac{\Delta DEO}{\Delta ABCO} = \frac{\frac{1}{2}\left(\frac{1}{2}\right)^2}{1} = \frac{1}{8}$$

Case 6 Flip a coin until a head appears

Step 1 Determine sample space

$$\Omega = \{H, TH, TTH, TTTH, TTTTH, ...\}$$

Step 2 Apply axioms

The most important axiom is:

$$P(\Omega) = P(H \cup TH \cup TTH \cup TTTH \cup TTTTH \cup ....) = 1$$

Since all the elements in the sample space are mutually exclusive, we have:

$$\begin{aligned} P(\Omega) &= P(H \cup TH \cup TTH \cup TTTH \cup TTTTH \cup ....) \\ &= P(H) + P(TH) + P(TTH) + P(TTTH) + P(TTTTH) + ... = 1 \end{aligned}$$

At first glance, it seems impossible to satisfy $P(\Omega) = 1$. We have infinite number of outcomes, yet the total probability needs to be one.

After more thoughts, however, we realize that it's easy to have $P(\Omega) = 1$. For example, if the probabilities follow a geometric progression, then we can have infinite # of probabilities, yet the sum of the total probability is finite. For example, if $P(H) = \frac{1}{2}$, $P(TH) = \frac{1}{2^2}$, $P(TTH) = \frac{1}{2^3}, \ldots$, then the total probabilities is

$$P(\Omega) = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + ... + \frac{1}{2^n} + ... = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1$$

Please note Case 6 is different from Case 4 and 5. Case 4 and 5 have infinite # of *continuous* outcomes. The only way to satisfy $P(\Omega) = 1$ is to have zero probability for each single outcome. This is why we need to have $P(x) = 0$ in Case 4 and $P(x, y) = 0$ in Case 5.

The outcomes in Case 6, however, are discrete. As a result, each outcome may have non-zero probability, yet the total probability can be finite.

Step 3 Assign probability

To assign probability for each outcome, we need to have detailed knowledge of the experiment. The experiment in this problem is "keep flipping coins until a head shows up." As you'll see later in this book, the probability model for experiment is a geometric distribution:

$$P(H)=p,\ P(TH)=p(1-p),\ P(TTH)=p(1-p)^2,\ldots$$
Where $0<p<1$

Then,

$$\begin{aligned}P(\Omega)&=P(H)+P(TH)+P(TTH)+P(TTTH)+P(TTTTH)+...\\&=p+p(1-p)+p(1-p)^2+p(1-p)^3+...\\&=\frac{p}{1-(1-p)}=1\end{aligned}$$

For now, don't worry about geometric distribution. Focus on understanding how to build a probability model.

Case 7 Flipping a tack

If we flip a tack many times, we want to find out how often we see the needle pointing up  or pointing down  .

Step 1 Determine sample space

$\Omega=\{U,D\}$, where U =up, D =down

Step 2 Apply axioms

$$P(U)\geq 0,\ P(D)\geq 0,\ P(\Omega)=P(U\cup D)=P(U)+P(D)=1$$

Step 3 Assign probability

It's reasonable to assume that outcomes U and D have fixed probabilities $P(U)=p$ and $P(D)=1-p$. However, unlike in tossing a fair coin where it's reasonable to assume $P(H)=P(T)$, here we can't assume $P(U)=P(D)$. We can't easily establish some sort of relationship between $P(U)$ and $P(D)$. What we can do, perhaps, is to flip a tack many times and determine the % of the times that the needle points up. However, this % is not $P(U)$; it's an estimate of $P(U)$.

Even though it's hard for us to assign probability to U and D, the axioms still hold. We'll always have

$$P(U) \geq 0,\ \ P(D) \geq 0,\ P(\Omega) = P(U \cup D) = P(U) + P(D) = 1$$

Hopefully by now you have an idea of how to build a framework to analyze a random process using set, sample space, and probability axioms.

Please note that when you solve problems, you don't have to formally list the 3 steps. As long as you can follow the essence of this 3-step process, you are fine.

Sample Problems and Solutions

Problem 1

John didn't bother learning the sample space or the probability axioms, yet he scored a 9 in Exam P. Explain this paradox.

Solution

This can happen. SOA exams mainly test a candidate's ability to apply theories to real world problems; SOA exams rarely test a concept for the sake of testing a concept. SOA never asked "Tell me what a sample space is." Instead, SOA asks you to solve a problem. If you can solve the problem, SOA assumes that you understand sample space and probability laws.

Though John didn't bother learning sample space or probability axioms, he understood the common sense behind sample space and axioms. For example, when calculating a probability, John always clearly specifies all of the possible outcomes, even though he doesn't know that the complete list of all the outcomes is called "sample space." John always assumes that the total probability for all the possible outcomes must add up to one, even though he doesn't' know that this is one of the axioms. So John has sufficient understanding of Exam P core concepts and deserves passing Exam P.

The key point to take home is this: when studying for Exam P and higher exams, focus on understanding the essence of a concept, not on memorizing jargon.

Problem 2

Let $A=\{1,3,7\}$, $U=\{1,2,3,4,5,6,7,8,9\}$, $B=\{4,7\}$.

Find

(1) All the subsets of A.
(2) $\overline{A}$.
(3) $A\cap B$.
(4) $A\cup B$.

Solution

(1) The subsets of A are as follows:

$$\varnothing, \{1\}, \{3\}, \{7\}, \{1,3\}, \{1,7\}, \{3,7\}, \{1,3,7\}$$

(2) $\overline{A}=\{2,4,5,6,8,9\}$

(3) $A\cap B=\{1,3,7\}\cap\{4,7\}=\{7\}$.

(4) $A\cup B=\{1,3,7\}\cup\{4,7\}=\{1,3,4,7\}$.

Problem 3

You toss a coin and record which side is up.

- Determine the sample space Ω.
- Determine the subset of Ω representing heads.
- Determine the subset of Ω representing tails.

Solution

If you toss a coin, you have a total of 2 possible outcomes: either heads or tails. As the result, the sample space is $\Omega=\{\text{H},\text{T}\}$.

The subset representing heads is $\{\text{H}\}$.

The subset representing tails is $\{\text{T}\}$.

Problem 4

You throw a die and record the outcomes.

- Determine the sample space Ω.
- Determine the subset of Ω representing an even number.
- Determine the subset of Ω representing an odd number.
- Determine the subset of Ω representing the smallest number.
- Determine the subset of Ω representing the largest number.
- Determine the subset of Ω representing a number no less than 3.

Solution

If you throw a die, you have a total of 6 possible outcomes -- you get 1,2,3,4, 5,or 6. As the result, the sample space is: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Within the sample space Ω, the subset representing an even number is $\{2, 4, 6\}$.

Within the sample space Ω, the subset representing an odd number is $\{1, 3, 5\}$.

Within the sample space Ω, the subset representing the smallest number is $\{1\}$.

Within the sample space Ω, the subset representing the biggest number is $\{6\}$.

Within the sample space Ω, the subset representing a number no less than 3 is $\{3, 4, 5, 6\}$.

Problem 5

An urn contains three balls – one red ball, one black ball, and one white ball. Two balls are randomly drawn from the urn and their colors are recorded.

Determine:

(1) The sample space.
(2) The event representing where a red ball and black ball are drawn.
(3) The probability that of the two randomly drawn balls, one is a red ball and the other a black ball.

Solution

Let R=red, B=black, and W=white.

The sample space is: { {R,B}, {B,R}, { R,W}, {W,R}, {B,W}, {W,B} }.

The event representing where a red ball and black ball are drawn is { {R,B}, {B,R} }.

The probability that of the two randomly drawn balls, one is a red ball and the other a black is:

$$\frac{\text{\# of elements in the event}}{\text{\# of elements in the sample space}} = \frac{2}{6} = \frac{1}{3}$$

In the above calculation, we can assume that each outcome in the sample space is equally likely to occur unless told otherwise.

Homework for you – redo all the problems listed in this chapter.

Chapter 4 Multiplication/addition rule, counting problems

Exam Advice

Counting problems can be extremely complex and difficult. However, previous SOA exam counting problems have been relatively simple and straightforward.

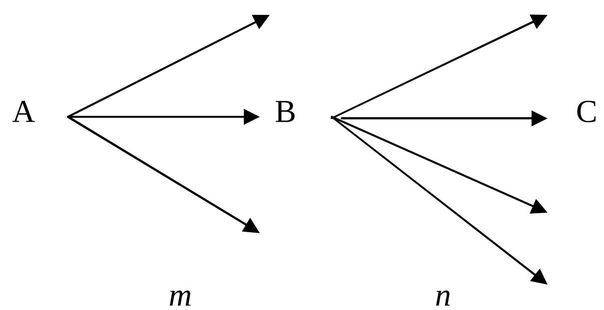
Focus on the basics. Don't spend too much time trying to master difficult problems.

Focus on the essence. Don't attempt to memorize the precise definition of terms such as "sampling with (or without) order and with (or without) replacement." As long as you know how to calculate the number of selections, it is not necessary to know whether a sampling method is with (or without) order and with (or without) replacement.

Multiplication Rule

You are going from City A to City C via City B. There are m different ways from City A to City B. There are n different ways from City B to City C. How many different ways are there from City A to City C?

There are $m \times n$ different ways from City A to City C.



Example. A department of a company has 3 managers, 10 professionals, and 5 secretaries. You want to choose one manager, one professional, and one secretary to form a committee. How many different ways can you form a committee?

Solution

There are 3 ways to choose a manager, 10 ways to choose a professional and 5 ways to choose a secretary. As a result, there are $3 \times 10 \times 5 = 150$ different ways to form a committee.

Addition Rule

If Event A and B are mutually exclusive, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

For any two events A and B, $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

If you understand the above formulas, then you have implicitly understood the addition rule without explicitly memorizing it.

Basic terms – Understand the essence. No need to memorize the definition.

Sample with order – The order by which items are listed matters. For example, two of the five equally qualified candidates (A, B, C, D, E) are chosen to fill two positions in a company --Vice President and Sales Manager. If we first list the candidate chosen for the Vice President position and then the candidate for the Sales Manager position, we can assume the order by which candidates are listed is important. Different orders represent different choices. For example, AB is different from BA. AB and BA two distinct entities.

Sample without order – The order by which items are listed does not matter. For example, two of the five equally qualified candidates (A, B, C, D, E) are chosen to form a committee. Evidently, a committee consisting of AB is the same committee consisting of BA. AB and BA are an identical entity.

Sample with replacement – After an item is taken out from the pool, it is immediately put back to the pool before the next random draw. As a result, the same item can be selected again and again. In addition, the number of items in the pool stays constant before each draw.

For example, you have 3 awards to give individually to 5 potential recipients. If one recipient is allowed to get multiple awards, then this is a sample with replacement.

Sample without replacement – Once an item is taken out of a pool, it permanently leaves the pool and can never be selected again. The number of items in the pool always decreases by one at the end of each random draw.
You have 3 awards to give individually to 5 potential recipients. If no recipients are allowed to get more than one award, then this is a sample without replacement.

Of course, you can sample with (or without) order and with (or without) replacement. These terms are explained below. Once again, when studying these terms, focus on learning the calculation, not the precise definition of the terms.

Select r out of n people to form a line (or select r out of n distinct objects) -- order matters and duplication is NOT allowed (or duplication is NOT possible) (called "sampling with order without replacement")

How many different ways can you form the line?

You can start filling the empty spots from left to right (you can also form the line from right to left and get the same result). You have n choices to fill the first spot, $(n-1)$ choices to fill the second spot,…, and $(n-r+1)$ choices to fill the r-th spot. Applying the multiplication rule, you have a total of

$$n\times(n-1)\times(n-2)\times...\times(n-r+1)=P_n^r$$

ways of forming the line.

If $r=n$, then we have $n!$ ways of forming the line.

Line up $n=r_1+r_2+...+r_k$ colored balls where r_1 balls have color 1 (such as red), r_2 balls have color 2 (such as black), …, r_k balls have color k

How many distinct ways are there to put n balls in a row? We have a total of $n!$ permutations if each ball is distinct. However, we have identical balls and we need to remove the permutations by these identical balls. So r_1 balls are identical and have $r_1!$ permutations; r_2 balls are identical and have $r_2!$ permutations. … r_k balls are identical and have $r_k!$ permutations. Thus, we need to divide $n!$ by $r_1!r_2!...r_k!$ to get the distinct number of permutations.

The total number of distinct permutations is:

$$\frac{n!}{r_1!r_2!...r_k!}=\frac{(r_1+r_2+...r_k)!}{r_1!r_2!...r_k!}$$

<u>Example.</u> You are using six letters A,A,A,B,B,C to form a six-letter symbol. How many distinct symbols can you create?

<u>Solution</u>

$$\frac{(3+2+1)!}{3!\,2!1!}=\frac{6!}{3!\,2!1!}=60$$

Select r out of n (where $n \geq r$) people to form a committee – order does NOT matter and duplication is NOT allowed (called "sampling without order and without replacement")

How many different ways can you form a committee?

First, you can choose r out of n people to form a line. You have a total of $n \times (n-1) \times (n-2) \times ... \times (n-r+1) = P_n^r$ ways to form a line with exactly r people standing in line. Out of these P_n^r ways, the same committee is repeated $r!$ times. As a result, you have a total of

$$\frac{P_n^r}{r!} = \frac{n!}{r!(n-r)!} = C_n^r$$

ways of forming a different committee.

C_n^r is often called the binomial coefficient.

If you have difficulty understanding why we need to divide P_n^r by $r!$, use a simple example. Say you are selecting two members out of three people A, B, C to form a committee. First, you choose 2 out of 3 people to stand in a line. You have a total of $P_3^2 = 3 \times 2 = 6$ ways of forming a 2-person line:

$$AB, BA, AC, CA, BC, CB$$

Of these 6 ways, you notice that the committee consisting of AB is the same as the committee consisting of BA; AC is the same as CA; BC is the same as CB. In other words, the number of committees is double counted. As a result, we need to divide P_3^2 by $2! = 2$ to get the correct number of committees that can be formed.

Form r-lettered symbols using n letters -- order matters and duplication is allowed (called "sampling with order and with replacement")

<u>Example</u>. Out of five letters a, b, c, d, e, you are forming three-lettered symbols such as aaa, aab,abc, …. How many different symbols can you build?

There is no relationship between r and n --- r can be greater than, less than, or equal to n.

You can use any of the n letters as the first letter of your symbol. You can use any of the n letters as the second letter of your symbol (because duplication is allowed). … You can use any of the n letters as the r-th letter of your symbol.

Applying the multiplication rule, we find that the number of unique symbols we can create is:

$$\underbrace{n \times n \times ... n}_{r} = n^r$$

Example. You are using four letters a, b, c, d to form a three-letter-symbol (for example, adb is one symbol). You are allowed to use the same letter multiple times when building a symbol. How many symbols can you form?

Solution

You can form a total of $4^3 = 64$ different symbols.

Collect a total of n items from r categories – order doesn't matter and duplication is allowed (called "sampling without order and with replacement")

This is best explained using an example. You want to collect a dozen silverware items from 3 categories: spoons, forks, and knives. You are not obligated to collect any items from any specific category, but you must collect a total of 12 items.

You have many choices. You can have 12 spoons, zero forks and zero knives; you can have 5 spoons, 5 forks, and 2 knives; you can have zero spoons, zero forks, and 12 knives. The only stipulation is that you must collect a total of 12 items.

How many different collections can you have?

To solve this problem, we need to use a diagram. We have a total of 14 empty boxes to fill. These 14 boxes consist of 12 silverware items we want to collect and 2 other boxes that indicate the ending of our selection of two categories.

□□□□□ □□□□□ □□□□

For example, we want to collect 5 spoons, 4 forks, and 3 knives. From left to right, we fill 5 empty boxes with spoons. In the sixth box, we put in an "X" sign indicating the end of our spoon collection. We then start filling the next 4 empty boxes with forks. Similarly, at the end of our fork collection, we fill an empty box with the "X" sign indicating the end of our folk collection. Finally, we fill the 3 remaining empty boxes with knives. However, this time we don't need any "X" sign to signal the end of our knife collection (our collection automatically ends there).

Because we have a total of 3 categories, we need to have 3-1=2 "X" signs.
This is why we have a total of 14 boxes. See diagram below.

S	S	S	S	S	X	F	F	F	F	X	K	K	K

Similarly, if we want to collect zero spoons, 8 forks, and 4 knives, the diagram will look like this:

X	F	F	F	F	F	F	F	F	X	K	K	K	K

If we want to collect zero spoons, zero forks, and 12 knives, the diagram will look like this:

X	X	K	K	K	K	K	K	K	K	K	K	K	K

Now we see that by positioning the two "X" boxes differently, we have different collections. We can put our two "X" boxes anywhere in the 14 available spots and each unique positioning represents one unique collection. Because we have a total of C_{14}^{2} different ways of positioning the two "X" boxes, we have a total of C_{14}^{2} different collections.

> Generally, we have $C_{n+r-1}^{r-1} = C_{n+r-1}^{n}$ different ways to assemble n items from r categories.

Example. Assume that a grocery store allows you to spend $5 to purchase a total of 10 items from a selection of four different vegetables: cucumbers, carrots, peppers, and tomatoes. Of these four categories, you can choose whatever items you want, but the total number of items you can have is 10. How many different vegetable combinations can you get with $5?

Solution

$$n = 10, r = 4, C_{n+r-1}^{r-1} = C_{13}^{3} = \frac{13 \times 12 \times 11}{3 \times 2 \times 1} = 286$$

You have 286 ways of buying 10 items.

SOA Problem (#1 Nov 2001)

An urn contains 10 balls; 4 red and 6 blue. A second urn contains 16 red balls and an unknown number of blue balls. A single ball is drawn from each urn. The probability that both balls are the same color is 0.44. Calculate the number of blue balls in the second urn.

$(A)\,4 \quad (B)\,20 \quad (C)\,24 \quad (D)\,44 \quad (E)\,64$

Solution

Let x =# of blue balls in the second urn.

	1st Urn	2^{nd} urn
Red ball	4	16
Blue ball	6	x
Total	10	x+16
Pr(red ball)	4/10	16/(16+x)
Pr(blue ball)	6/10	x/(16+x)

Pr(both red) + Pr(both blue)=0.44

Because the first urn and second urn are independent,

Pr(both red)= $\frac{4}{10}\times\frac{16}{16+x}$, Pr(both blue)= $\frac{6}{10}\times\frac{x}{16+x}$

So we have

$$\frac{4}{10}\times\frac{16}{16+x}+\frac{6}{10}\times\frac{x}{16+x}=0.44 \quad \Rightarrow x=4$$

Homework for you – redo all the problems listed in this chapter.

Chapter 5 Probability laws and "whodunit"

In virtually every exam sitting, there are several probability problems which require you to "sort out who did what." These problems are more a test of common sense than a test of one's profound understanding of probability concepts.

This is a simple set-up of such a problem:

> In a group of people, X% did A, Y% did B, Z% did both A and B. What percentage did A but not B?

This problem is simple because you have only two basic categories (A and B) and $2^2=4$ categories ($AB, \bar{A}B, A\bar{B}, \bar{A}\bar{B}$) to divide the population (AB =doing A and B, $\bar{A}B$ =doing B but not A, $A\bar{B}$ =doing A but not B, and $\bar{A}\bar{B}$ =doing neither A nor B). It is easy to mentally track four possibilities.

A more complex problem requires you to sort people or things under three categories. The problem is set up something like this:

> Of a group of people, X% did A, Y% did B, Z% did C, M% did A and B, N% did B and C, etc. What percentage of the people did A and B and C? Or did nothing (no A, no B, and no C)? Or did some other combinations of A, B, and C?

This problem is much harder. Essentially, you have a total of $2^3=8$ possible categories:

$$ABC,\quad A\bar{B}C,\quad AB\bar{C},\quad A\bar{B}\bar{C},\quad \bar{A}BC,\quad \bar{A}\bar{B}C,\quad \bar{A}B\bar{C},\quad \bar{A}\bar{B}\bar{C}$$

It is hard to mentally track 8 possibilities, not only under exam conditions where you have about 3 minutes to solve a problem, but even under conditions where time is not a big constraint. To solve this kind of problem right, you need a way to track the different possibilities.

Is it likely to have an exam question where you have 4 basic categories (A, B, C, and D) and a total of $4^2=16$ possible combinations for you to sort out? Not very likely. It takes too much work for you to solve it. It takes too much work for SOA to write this question and come up with an answer too. However, if this kind of difficult problem does show up, you will probably want to skip it and focus on easier problems, unless your goal is to score a 10.

How to tackle the problem in 3 minutes

There are 2 approaches to "sort out who did what" – the formula-driven approach and the common sense approach. You might want to familiarize yourself with both approaches and choose one that you like better.

Formula-driven approach

There are a series of formulas you can use to solve the "who did what" problem. Some of the formulas are easier to memorize than others. I recommend that you memorize just the basic formulas – they are more useful and versatile than complex formulas.

For complex formulas, don't worry about memorizing them. Instead, get a feel for their logic.

Basic formulas you need to memorize:

$P(A)+P(\overline{A})=1$ (some textbooks use A^c for $\overline{A}$)

$\mathrm{P}(A\cup B)=\mathrm{P}(A)+\mathrm{P}(B)-\mathrm{P}(A\cap B)$
In other words, $P(A \text{ or } B)=P(A)+P(B)-P(A \text{ and } B)$

On the right hand side of the above formula, we have to subtract $P(A \text{ and } B)$ because it's double counted in both $\mathrm{P}(A)$ and $\mathrm{P}(B)$.

$P(A\cap B)=P(A)P(B|A)$
Intuitively, $P(A\cap B)$ = % of population who did both A and B
$P(A)$ = % of the population who did A
$P(B|A)$ =% of the people who did A did B

$P(A\cap B)=P(B)P(A|B)$ says that
the % of population who did both A and B
= % of the population who did B $\times$ % of the people who did B did A
This makes good sense.

$P(A\cap B)=P(A)P(B)$ if A and B are independent

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$$
The proof is simple.

$$P(A\cup B\cup C)=P\left[A\cup(B\cup C)\right]=P(A)+P(B\cup C)-P\left[A\cap(B\cup C)\right]$$

$$P(B\cup C)=P(B)+P(C)-P(B\cap C)$$

$$P\left[A\cap\left(B\cup C\right)\right]=P\left[\left(A\cap B\right)\cup\left(A\cap C\right)\right]$$
$$=P\left(A\cap B\right)+P\left(A\cap C\right)-P\left[\left(A\cap B\right)\cap\left(A\cap C\right)\right]$$

$$P\left[\left(A\cap B\right)\cap\left(A\cap C\right)\right]=P\left(A\cap B\cap C\right)$$

Putting everything together, we have:

$$P\left(A\cup B\cup C\right)=P\left(A\right)+P\left(B\right)+P\left(C\right)-P\left(A\cap B\right)-P\left(A\cap C\right)-P\left(B\cap C\right)+P\left(A\cap B\cap C\right)$$

The above formula is ugly but SOA likes to test it. So memorize it.

Complex formulas --- don't memorize them, but do have a feel for their logic:

Associate Laws:

$A\cup(B\cup C)=(A\cup B)\cup C$

In other words, A or (B or C)=(A or B) or C. This says that if you want to count the % of people who did at least one of the three tasks -- A, B, and C, you can count it in two ways. The first counting method is to compile two lists of people – one lists those who did at least one of the two tasks, B and C, and the other lists those who did Task A. If you merge these two lists, you should get the total % of people who did at least one task.

The second counting method is to come up with the list of people who did at least one of the two tasks A and B and the list of people who did C. Then you merge these two lists.

Intuitively, you can see that the two counting methods should generate the identical result.

$A\cap(B\cap C)=(A\cap B)\cap C$

In other words, A and (B and C)=(A and B) and C. The left-hand side and the right-hand side are two ways of counting who did all three tasks – A, B, and C. The two counting methods should give identical results.

Communicative Laws:

$A\cup B=B\cup A$ (In other words, A or B=B or A. Makes intuitive sense.)

$A\cap B=B\cap A$ (In other words, A and B=B and A. Makes intuitive sense.)

Distributive Laws:

$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

In other words, (A or B) and C=(A and C) or (B and C)

$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

In other words, (A and B) or C=(A or C) and (B or C)

Distributive Laws are less intuitive. You might want to use some concrete examples to convince yourself that the laws are true.

DeMorgan's Laws:

$$\overline{A_1 \cup A_2 \cup A_3 \cup ... \cup A_n} = \overline{A_1} \cap \overline{A_2} \cap \overline{A_3} \cap ... \cap \overline{A_n}$$

In other words, Not any of A_1 or A_2 or A_3 or … A_n is the same as No A_1 and No A_2 and No A_3 and … No A_n .They are just two ways of counting the % of people who did not do any of the n tasks.

$$\overline{\mathrm{A}_1 \cap \mathrm{A}_2 \cap \mathrm{A}_3 \cap \ldots \cap \mathrm{A}_n} = \overline{\mathrm{A}_1} \cup \overline{\mathrm{A}_2} \cup \overline{\mathrm{A}_3} \cup \ldots \cup \overline{\mathrm{A}_n}$$

The left-hand side and the right-hand are two ways of counting the % of people who didn't complete all of the n tasks (i.e. who did not do any task at all, or who did some but not all tasks.)

Common Sense Approach

If you dislike memorizing formulas, you can just use your common sense to solve the problem. However, you will still need some tools (a table or a diagram called a Venn diagram) to help you sort out who did what. I will show you how.

Sample Problems and Solutions

Problem 1 (#1 May 2000)

The probability that a visit to a primary care physician's (PCP) office results in neither lab work nor referral to a specialist is 35%. Of those coming to a PCP's office, 30% are referred to specialists and 40% require lab work.
Determine the probability that a visit to a PCP's office results in both lab work and referral to a specialist.

A 0.05 B 0.12 C 0.18 D 0.25 E 0.35

Solution

Formula-driven approach

Let L=lab work and R=referral to specialists. We need to find $P(L \cap R)$.
We are given: $P(L) = 40\%$, $P(R) = 30\%$, "neither R nor L Neither R=35%."

The tricky part is "neither R nor L Neither R=35%." Generally,

Neither A nor B = (No A) and (No B)= $\overline{A} \cap \overline{B} = \overline{A \cup B}$
You might want to memorize this.

Once you understand "neither …nor…," the rest is really simple.

$P\left(\overline{R \cup L}\right)$=35%, $\Rightarrow$ $P(R \cup L)$=1- $P\left(\overline{R \cup L}\right)$=65%.

$P(R \cup L) = P(R) + P(L) - P(R \cap L)$
65% = 40% + 30% - $P(R \cap L)$
$P(R \cap L)$=5%

Common sense approach—table method

You can come up with a table that looks like this:

	A	B	C	D
1		Referral	No referral	Sum
2	Lab work	?		40%
3	No lab work		35%	
4	Sum	30%		100%

Cell(D,2)=% of visits leading to lab work=40%
Cell(B,4)=% of visits leading to referral=30%
Cell(C,3)=% of visits leading to no lab work and no referral=35%

We need to find Cell(B,2)=% of visits leading to both lab work and referral.

Because Pr(Referral)+Pr(no referral)=1, Pr(lab work)+Pr(no lab work)=1,

Cell(D,4)=100%=Cell(D,2)+Cell(D,3)=Cell(B,4)+Cell(C,4)

We update the table with the above info:

	A	B	C	D
1		Referral	No referral	Sum
2	Lab work	?		40%
3	No lab work		35%	60%
4	Sum	30%	70%	100%

Because Cell(B,3)+Cell(C,3)=60%, Cell(B,3)=25%
Because Cell(B,2)+Cell(B,3)=30%, Cell(B,2)=5%.

The final table looks like this:

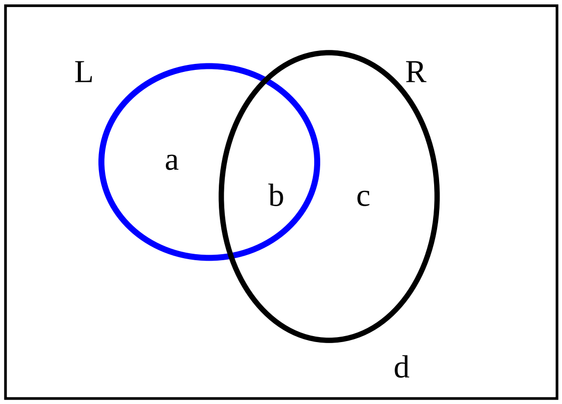
	A	B	C	D
1		Referral	No referral	Sum
2	Lab work	5%	35%	40%
3	No lab work	25%	35%	60%
4	Sum	30%	70%	100%

You can verify that the numbers in the above table satisfy the following relationships:

Cell(B,2)+Cell(C,2)=Cell(D,2)
Cell(B,3)+Cell(C,3)=Cell(D,3)
Cell(B,2)+Cell(B,3)=Cell(B,4)
Cell(C,2)+Cell(C,3)=Cell(C,4)

Common sense approach – Venn diagram

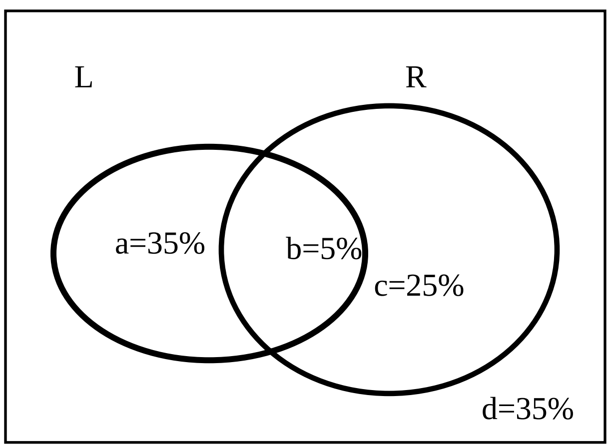
First draw a graph:



L=a+b=40%
R=b+c=30%
d=35%

a+b+c+d=100%

You can easily find all the variables by solving the above equations.
The result: a=35%, b=5%, c=25%



You can also see from the above diagram:

$$a = P\left(L \cap \overline{R}\right), \qquad b = P\left(L \cap R\right)$$
$$c = P\left(\overline{L} \cap R\right), \qquad d = P\left(\overline{L} \cap \overline{R}\right)$$

Problem 2 (#5 May 2003)

An insurance company examines its pool of auto insurance customers and gathers the following information:

(i) All customers insure at least one car.
(ii) 70% of the customers insure more than one car.
(iii) 20% of the customers insure a sports car.
(iv) Of the customers who insure more than one car, 15% insure a sports car.

Calculate the probability that a randomly selected customer insures exactly one car and that this car is not a sports car.

Solution

Formula-driven approach
Let C=insuring more than one car, S=insuring a sports car.

All customers insure at least one car $\Rightarrow$ $\overline{C}$=insuring exactly one car.

$\Rightarrow$ insuring exactly one car and this car is not a sports car = $\overline{C} \cap \overline{S}$

So we need to find out $P\left(\overline{C} \cap \overline{S}\right)$. Since $P\left(\overline{C} \cap \overline{S}\right)$ has two negation operators, we might want to simplify it first:

$$\overline{C} \cap \overline{S} = 1 - \overline{\overline{C} \cap \overline{S}} = 1 - C \cup S \quad \text{(DeMorgan's law)}$$

Intuitively, $\overline{C} \cap \overline{S}$ means no C and no S. $C \cup S$ means C or S or both. This is why $\overline{C} \cap \overline{S} = 1 - C \cup S$.

$$\Rightarrow \quad P(C \cup S) = P(C) + P(S) - P(C \cap S) = P(C) + P(S) - P(C)P(S|C)$$

$P(C)$=70%, $P(S)$=20%, and $P(S|C)$=15%.

$$\Rightarrow \quad P(C \cup S) = P(C) + P(S) - P(C)P(S|C)$$
$$=70\% + 20\% - 70\% \times 15\% = 79.5\%$$

$$\Rightarrow \quad P\left(\overline{C} \cap \overline{S}\right) = 1 - P(C \cup S) = 1 - 79.5\% = 20.5\%$$

Common sense approach – table method

Let's develop the following table to sort out who did what.

	A	B	C	D	E
1		Insure zero sports car	Insure one sports car	Insure 2 or more sports car	Sum
2	Insure one car			0%	30%
3	Insure more than one car				70%
4	Sum		20%		100%

Cell(C,4)=*P*(insuring one sports car)=20%
Cell(E,3)=*P*(insuring more than one car)=70%
Cell(E,2)=*P*(insuring one car)=1–70%=30%
Cell(D,2)=0 (if you insure only car, you cannot insure two or more sports cars)

Let's fill out some of the remaining cells.

Cell(C,3)=70%(15%)=10.5% (15% of those who insured more than one car insured one sports car.)

Cell(C,2)=Cell(C,4) –Cell(C,3)=20% – 10.5%=9.5%

Cell(B,2)=Cell(E,2) –[Cell(C,2) + Cell(D,2)]
= 30% - [9.5% + 0] = 20.5%

The updated table:

	A	B	C	D	E
1		Insure zero sports car	Insure one sports car	Insure 2 or more sports car	Sum
2	Insure one car	**20.5%**	9.5%	0%	30%
3	Insure more than one car		70%(15%) =10.5%		70%
4	Sum		20%		100%

Problem 3

A survey of a group of consumers finds the following result:

59% bought life insurance products.
34% bought annuity products.
48% bought mutual fund products.
14% bought none (no life insurance, no annuity, no mutual fund)
17% bought both life insurance and annuity products.
30% bought both life insurance and mutual fund products.
19% bought both annuity and mutual fund products.

Find the % of the group who bought all three products (life insurance, annuities, and mutual funds).

Solution

Formula-driven approach
Let L,A, and M stand for that a consumer surveyed owns life insurance products, annuity products, and mutual fund products respectively.

$$P(L)=59\%,\ P(A)=34\%,\ P(M)=48\%$$
$$P(L\cap A)=17\%,\ P(L\cap M)=30\%,\ P(A\cap M)=19\%$$
$$P\left(\overline{L}\cap\overline{A}\cap\overline{M}\right)=14\% \quad \text{(14\% bought none)}$$

$$\Rightarrow\ P(L\cup A\cup M)=1-P\left(\overline{L\cup A\cup M}\right)=1-P\left(\overline{L}\cap\overline{A}\cap\overline{M}\right)=1\text{-}14\%=86\%$$

We can also derive the above formula by reasoning. 14% bought none (no life insurance, no annuity, no mutual fund). The percentage of customers who bought at least one product is $P(L\cup A\cup M)$=1-14%=86%

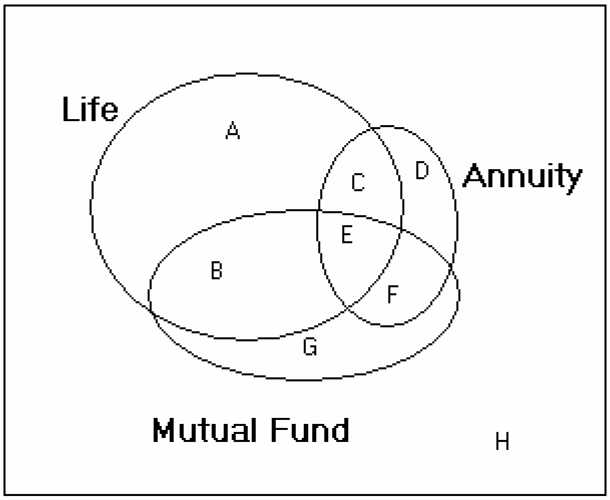
Plug in the above data into a memorized formula:
$$P(L\cup A\cup M)$$
$$=P(L)+P(A)+P(M)-P(L\cap A)-P(L\cap M)-P(A\cap M)+P(L\cap A\cap M)$$

86%=59%+34%+48%-17%-30%-19% $+P(L\cap A\cap M)$

Then $P(L\cap A\cap M)$=11%

Venn Diagram:

Draw a Venn diagram as below.



We will keep track of all the components in the above diagram (letters A through H):

A=Life only (no Annuity, no Mutual Fund)
D=Annuity only (no Life, no Mutual Fund)
G=Mutual Fund only (no Life, no Annuity)

E=Life and Annuity and Mutual Fund
C=Life and Annuity only, no Mutual Fund
B=Life and Mutual Fund only, no Annuity
F=Mutual Fund and Annuity Only, no Life

C+E=Life and Annuity (whether there's Mutual Fund or not)
B+E=Life and Mutual Fund (whether there's Annuity or not)
F+E=Mutual Fund and Annuity (whether there's Life or not)

H=Nothing (i.e. no Life, no Annuity, no Mutual Fund)

Life=A+B+C+E (sum of all the Life purchases)
Annuity=C+D+E+F (sum of all the Annuity purchases)
Mutual Fund=B+E+F+G (sum of all the Mutual Fund purchases)

Life + Annuity + Mutual Fund + Nothing = 100%

We are asked to find E.
Next, we use the following information to set equations:

59% bought life insurance products. $\Rightarrow$ A+B+C+E=59% (1)
34% bought annuity products. $\Rightarrow$ C+D+E+F=34% (2)
48% bought mutual fund products. $\Rightarrow$ B+E+F+G=48% (3)

14% bought nothing $\Rightarrow$ H=14% (4)

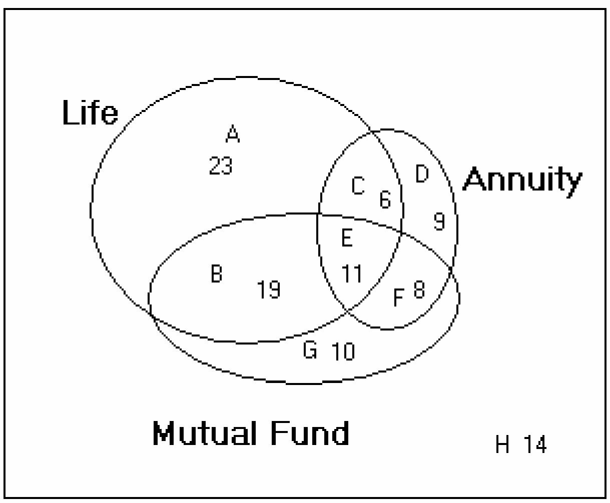
17% bought both life insurance and annuity products. $\Rightarrow$C+E=17% (5)

30% bought both life insurance and mutual fund products. $\Rightarrow$B+E=30% (6)

19% bought both annuity and mutual fund products. $\Rightarrow$F+E=19% (7)

Total is 100% $\Rightarrow$A+B+C+D+E+F+G+H=100% (8)

We have 8 variables A through H. We have 8 equations. If you have lot of patience, you should be able to solve these equations and track each component as follows:



So the % of people who purchase every product is E=11%.

In contrast to the formula-driven approach, the Venn Diagram method does not require you to memorize formulas. I recommend that you get comfortable with both methods.

Please note that the table approach is too cumbersome to track things if you have 3 categories (like L, A, M in this problem) or more.

Homework for you: #1 May 2000; #9 Nov 2001; #3 Nov 2000; #31 May 2001; #1 May 2003

Chapter 6 Conditional Probability

Often times we have partial information about an experiment. This partial information will change our calculation of probability.

Example 1. We throw two dies and sum up the two numbers. We want to find out the probability that the sum is 8. The sample space is:

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

Of the above 36 outcomes, 5 outcomes (blue numbers) give us a sum of 8. So the probability of getting a total of 8 is 5/36.

Now suppose we know that the throw of the 1st die gives a 5. Given this information, what's the probability of still getting a sum of 8? Given the 1st number is 5, the total outcomes are listed below in red. Of these 6 outcomes, only (5,3) gives us an 8. So the probability of getting an 8 given the 1st number is 5 is 1/6, not 5/36. Knowing that "the 1st die gives us a 5" altered the probability.

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

Why does the new information alter the probability? Because the new information alters the sample space.

The sample space before we know "the 1st die gives us a 5" is:

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

The sample space after we know "the 1st die gives us a 5" is:

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

The new sample space is smaller than the original one.

The arrival of new information reduces our sample space. Consequently, we have to recalculate probability in the reduced sample space.

Let's look at the formula for conditional probability.

$$P(B|A)=\frac{\text{\# of elements in } A\cap B}{\text{\# of elements in the reduced sample space } A}$$

$$=\frac{\text{\# of elements in } A\cap B}{\text{\# of elements in the original sample space } \Omega}\div\frac{\text{\# of elements in the reduced sample space of } A}{\text{\# of elements in the original sample space } \Omega}$$

But

$$\frac{\text{\# of elements in } A\cap B}{\text{\# of elements in the original sample space } \Omega}=P(A\cap B)$$

$$\frac{\text{\# of elements in the reduced sample space of } A}{\text{\# of elements in the original sample space } \Omega}=P(A)$$

$$\Rightarrow \quad P(B|A)=\frac{P(A\cap B)}{P(A)}$$

In this example, let

A =getting a 5 if we throw the 1st die (new information)

B =getting a sum of 8 if we throw two dies.

Then $P(B|A)$ is the probability of getting a sum of 8 if the 1st die is 5.

In this example,

$$P(A)=\frac{1}{6}.$$

$$P(A\cap B)=P(\text{1st \# is 5; 2nd \# is 3})=P(\text{1st \# is 5})\times P(\text{2nd \# is 3})=\frac{1}{6}\times\frac{1}{6}$$

$$P(B|A)=\frac{P(A\cap B)}{P(A)}=\frac{\frac{1}{6}\times\frac{1}{6}}{\frac{1}{6}}=\frac{1}{6}$$

Please note that all the probability formulas that hold in the original sample space still hold in the reduced sample space.

For example:

$$P(B|A)+P(\overline{B}|A)=1$$
$$P(B\cup C|A)=P(B|A)+P(C|A)-P(B\cap C|A)$$

Example 2. We throw a die. Let X represent the # that is facing up. Calculate $P(X=5|X>4)$.

Solution

$$P(X=5|X>4)=\frac{P(X=5\cap X>4)}{P(X>4)}=\frac{P(X=5)}{P(X>4)}=\frac{P(X=5)}{P(X=5\cup X=6)}=\frac{1}{2}$$

Let's look at how the sample space is altered. The original sample space:

1	2	3	4	5	6

The new sample space after we know that $X>4$:

5	6

The probability of getting a 5 in the new sample space is 0.5.

Example 3. (#39, May 2001) An insurance company insures a large # of homes. The insured value, X , of a randomly selected home is assumed to follow a distribution with density function:

$$f(x)=\begin{cases}3x^{-4} & \text{for } x>1\\ 0 & \text{elsewhere}\end{cases}$$

Given that a randomly selected home is insured for at least 1.5, what is the probability that it is insured for less than 2?

Solution

$$P\left(X \leq 2 \mid X > 1.5\right) = \frac{P\left(X \leq 2 \cap X > 1.5\right)}{P\left(X > 1.5\right)} = \frac{P\left(1.5 < X \leq 2\right)}{P\left(X > 1.5\right)}$$

$$P\left(1.5 < X \leq 2\right) = \int_{1.5}^{2} f\left(x\right)dx = \int_{1.5}^{2} 3x^{-4}dx = -x^{-3}\Big|_{1.5}^{2} = 1.5^{-3} - 2^{-3}$$

$$P\left(1.5 < X\right) = \int_{1.5}^{+\infty} f\left(x\right)dx = \int_{1.5}^{+\infty} 3x^{-4}dx = -x^{-3}\Big|_{1.5}^{+\infty} = 1.5^{-3}$$

$$P\left(X \leq 2 \mid X > 1.5\right) = \frac{P\left(1.5 < X \leq 2\right)}{P\left(X > 1.5\right)} = \frac{1.5^{-3} - 2^{-3}}{1.5^{-3}} = 0.5781$$

Homework for you:

#7, Nov 2000; #17, #32, Nov 2001.

Chapter 7 Bayes' theorem and posterior probabilities

Prior probability. Before anything happens, as our baseline analysis, we believe (based on existing information we have up to now or using purely subjective judgment) that our total risk pool consists of several homogenous groups. As a part of our baseline analysis, we also assume that these homogenous groups have different sizes. For any insured person randomly chosen from the population, he is charged a weighed average premium.

As an over-simplified example, we can divide, by the aggressiveness of a person's driving habits, all insured's into two homogenous groups: aggressive drivers and non-aggressive drivers. In regards to the sizes of these two groups, we assume (based on existing information we have up to now or using purely subjective judgment) that the aggressive insured's account for 40% of the total insured's and non-aggressive account for the remaining 60%.

So for an average driver randomly chosen from the population, we charge a weighed average premium rate (we believe that an average driver has some aggressiveness and some non-aggressiveness):

Premium charged on a person randomly chosen from the population = 40%*premium rate for an aggressive driver's rate + 60%*premium rate for a non-aggressive driver's rate

Posterior probability. Then after a year, an event changed our belief about the makeup of the homogeneous groups for a specific insured. For example, we found in one year one particular insured had three car accidents while an average driver had only one accident in the same time period. So the three-accident insured definitely involved more risk than did the average driver randomly chosen from the population. As a result, the premium rate for the three-accident insured should be higher than an average driver's premium rate.

The new premium rate we will charge is still a weighted average of the rates for the two homogeneous groups, except that we use a higher weighting factor for an aggressive driver's rate and a lower weighting factor for a non-aggressive driver's rate.

For example, we can charge the following new premium rate:

Premium rate for a driver who had 3 accidents last year = 67%* premium rate for an aggressive driver's rate + 33%* premium rate for a non-aggressive driver's rate

In other words, we still think this particular driver's risk consists of two risk groups – aggressive and non-aggressive, but we alter the sizes of these two risk groups for this specific insured. So instead of assuming that this person's risk consists of 40% of an aggressive driver's risk and 60% of a non-aggressive driver's risk, we assume that his risk consists of 67% of an aggressive driver's risk and 33% of a non-aggressive driver's risk.

How do we come up with the new group sizes (or the new weighting factors)? There is a specific formula for calculating the new group sizes:

For any given group,

Group size after an event
=K × the group size before the event × this group's probability to make the event happen.

K is a scaling factor to make the sum of the new sizes for all groups equal to 100%.

In our example above, this is how we got the new size for the aggressive group and the new size for the non-aggressive group. Suppose we know that the probability for an aggressive driver to have 3 car accidents in a year is 15%; the probability for a non-aggressive driver to have 3 car accidents in a year is 5%. Then for the driver who has 3 accidents in a year,

the size of the aggressive risk for someone who had 3 accidents in a year
= K l (prior size of pure aggressive risk)
  l (probability of an aggressive driver having 3 car accidents in a year)
= K (40%)(15%)

the size of the non-aggressive risk for someone who had 3 accidents in a year
= K l (prior size of the non-aggressive risk)
  l (probability of a no- aggressive driver having 3 car accidents in a year)
= K (60%) (5%)
K is a scaling factor such that the sum of posterior sizes is equal to one. So

K (40%) (15%) + K (60%) (5%) =1,

$$K = \frac{1}{40\%(15\%) + 60\%(5\%)} = 11.11\%$$

the size of the aggressive risk for someone who had 3 accidents in a year
= 11.11% (40%) (15%)= 66.67%

the size of the non-aggressive risk for someone who had 3 accidents in a year
=11.11% (60%) (5%) = 33.33%

The above logic should make intuitive sense. The bigger the size of the group prior to the event, the higher contribution this group will make to the event's occurrence; the bigger the probability for this group to make the event happen, the higher the contribution this group will make to the event's occurrence. So the product of the prior size of the group and the group's probability to make the event happen captures this group's total contribution to the event's occurrence.

If we assign the post-event size of a group proportional to the product of the prior size and the group's probability to make the event happen, we are really assigning the post-event size of a group proportional to this group's total contribution to the event's occurrence. Again, this should make sense.

Let's summarize the logic for finding the new size of each group in the following table:

Event: An insured had 3 accidents in a year.

A	B	C	D=(scaling factor K) ×B×C
Homogenous groups (which are 2 components of a risk)	Before-event group size	Group's probability to make the even happen	Post-event group size
Aggressive	40%	15%	K×40%×15% $=\dfrac{40\%\times 15\%}{40\%\times 15\%+60\%\times 5\%}$
Non-aggressive	60%	5%	K×60%×5% $=\dfrac{60\%\times 5\%}{40\%\times 15\%+60\%\times 5\%}$

We can translate the above rule into a formal theorem:

If we divide the population into *n* non-overlapping groups $G_1, G_2, \ldots, G_n$ such that each element in the population belongs to one and only one group, then after the event *E* occurs,

$$\Pr(G_i \mid E) = K \times \Pr(G_i) \times \Pr(E \mid G_i)$$

K is a scaling factor such at

$$K \times [\Pr(G_1 \mid E) + \Pr(G_2 \mid E) + \ldots + \Pr(G_n \mid E)] = 1$$

Or $K \times [\Pr(G_1) \times \Pr(E \mid G_1) + \Pr(G_2) \times \Pr(E \mid G_2) + \ldots + \Pr(G_n) \times \Pr(E \mid G_n)] = 1$

So $K = \dfrac{1}{\Pr(G_1) \times \Pr(E \mid G_1) + \Pr(G_2) \times \Pr(E \mid G_2) + \ldots + \Pr(G_n) \times \Pr(E \mid G_n)}$

And $\Pr(G_i \mid E) = \dfrac{\Pr(G_i) \times \Pr(E \mid G_i)}{\Pr(G_1) \times \Pr(E \mid G_1) + \Pr(G_2) \times \Pr(E \mid G_2) + \ldots + \Pr(G_n) \times \Pr(E \mid G_n)}$

$\Pr(G_i \mid E)$ is the conditional probability that G_i will happen given the event *E* happened, so it is called the posterior probability. $\Pr(G_i \mid E)$ can be conveniently interpreted as the new size of Group G_i after the event *E* happened. Intuitively, probability can often be interpreted as a group size.

For example, if a probability for a female to pass Course 4 is 55% and male 45%, we can say that the total pool of the passing candidates consists of 2 groups, female and male with their respective sizes of 55% and 45%.

$\Pr(G_i)$ is the probability that G_i will happen prior to the event *E*'s occurrence, so it's called prior probability. $\Pr(G_i)$ can be conveniently interpreted as the size of group G_i prior to the occurrence of *E*.

$\Pr(E \mid G_i)$ is the conditional probability that *E* will happen given G_i has happened. It is the Group G_i's probability of making the event E happen. For example, say a candidate who has passed Course 3 has 50% chance of passing Course 4, that is to say:

Pr(passing Course 4 / passing Course 3)=50%

We can say that the people who passed Course 3 have a 50% of chance of passing Course 4.

How to tackle the problem in 3 minutes

Solving a posterior probability problem for a discrete random variable is a simple matter of applying Bayes' theorem to a specific situation. Remember that for Bayes' theorem to work, you first need to divide the population into several non-overlapping groups such that everybody in the population belongs to one and only one group.

Sample Problems and Solutions

Before we jump into the formula, let's look at a sixth-grade level math problem, which requires zero knowledge about probability. If you understand this problem, you should have no trouble understanding Bayes' Theorem.

Problem 1

A rock is found to contain gold. It has 3 layers, each with a different density of gold. You are given:

- The top layer, which accounts for 80% of the mass of the rock, has a gold density of only 0.1% (i.e. the amount of gold contained in the top layer is equal to 0.1% of the mass of the top layer).

- The middle layer, which accounts for 15% of the rock's mass, has a gold density of 0.05%.

- The bottom layer, which accounts for only 5% of the rock's mass, has a gold density of 0.002%.

Questions

What is the rock's density of gold (i.e.: what % of the rock's mass is gold)?

Of the total amount of gold contained in the rock, what % of gold comes from the top layer? What % from the middle layer? What % comes from the bottom layer?

Solution

Let's set up a table to solve the problem. Assume that the mass of the rock is one (can be 1 pound, 1 gram, 1 ton – it doesn't matter).

	A	B	C	D=B×C	E=D/0.000876
1	Layer	Mass of the layer	Density of gold in the layer	Mass of gold contained in the layer	Of the total amount of gold in the rock, what % comes from this layer?
2	Top	0.80	0.100%	0.000800	91.3%
3	Middle	0.15	0.050%	0.000075	8.6%
4	Bottom	0.05	0.002%	0.000001	0.1%
5	Total	1.00		0.000876	100%

As an example of the calculations in the above table,

Cell(D,2)=0.8×0.100%=0.000800,
Cell(D,5)=0.000800+0.000075+0.000001=0.000876,
Cell(E,2)= 0.000800/0.000876=91.3%.

So the rock has a gold density of 0.000876 (i.e. 0.0876% of the mass of the rock is gold).

Of the total amount of gold contained in the rock, 91.3% of the gold comes from the top layer, 8.6% of the gold comes from the middle layer, and the remaining 0.1% of the gold comes from the bottom layers. In other words, the top layer contributes to 91.3% of the gold in the rock, the middle layer 8.6%, and the bottom layer 0.1%.

The logic behind this simple math problem is exactly the same logic behind Bayes' Theorem.

Now let's change the problem into one about prior and posterior probabilities.

Problem 2

In underwriting life insurance applications for nonsmokers, an insurance company believes that there's an 80% chance that an applicant for life insurance qualifies for the standard nonsmoker class (which has the standard underwriting criteria and the standard premium rate); there's a 15% chance that an applicant qualifies for the preferred smoker class (which has more stringent qualifying standards and a lower premium rate than the standard nonsmoker class); and there's a 5% chance that the applicant qualifies for the super preferred class (which has the highest underwriting standards and the lowest premium rate among nonsmokers).

According to medical statistics, different nonsmoker classes have different probabilities of having a specific heart-related illness:

- The standard nonsmoker class has 0.100% of chance of getting the specific heart disease.
- The preferred nonsmoker class has 0.050% of chance of getting the specific heart disease.
- The super preferred nonsmoker class has 0.002% of chance of getting the specific heart disease.

If a nonsmoking applicant was found to have this specific heart-related illness, what is the probability of this applicant coming from the standard risk class? What is the probability of this applicant coming from the preferred risk class? What is the probability of this applicant coming from the super preferred risk class?

Solution

The solution to this problem is exactly the same as the one to the rock problem.

Event: the applicant was found to have the specific heart disease

	A	B	C	D=B×C	E=D/0.000876 (i.e. the scaling factor =1/0.000876)
1	Group	Before-event size of the group	This group's probability of having the specific heart illness	After-event size of the group (not yet scaled)	After-event size of the group (scaled)
2	Standard	0.80	0.100%	0.000800	91.3%

3	Preferred	0.15	0.050%	0.000075	8.6%
4	Super Preferred	0.05	0.002%	0.000001	0.1%
5	Total	1.00		0.000876	100%

So if the applicant was found to have the specific heart disease, then

There's a 91.3% chance he comes from the standard risk class;
There's an 8.6% chance he comes from the preferred risk class;
There's a 0.1% chance he comes from the super preferred risk class.

Problem 3 (continuous random variable)

You are tossing a coin. Not knowing p, the success rate of a heads showing up in one toss of the coin, you subjectively assume that p is uniformly distributed over $[0,1]$. Next, you do an experiment by tossing the coin 3 times. You find that, in this experiment, 2 out of 3 tosses have heads.

Calculate the posterior probability p.

Solution

Event: getting 2 heads out of 3 tosses.

	A	B	C	D=B×C	E=D× Scaling factor
1	Group	Before-event size of the group	This group's probability to make the event happen	After-event size of the group (not yet scaled)	After-event size of the group (scaled)
2	Any p in [0,1]	1	$C_3^2 p^2 (1-p)$	$C_3^2 p^2 (1-p)$	$\dfrac{C_3^2 p^2 (1-p)}{\int_0^1 C_3^2 p^2 (1-p) dp}$
3	Total	1		$\int_0^1 C_3^2 p^2 (1-p) dp$	100%

The key to solving this problem is to understand that we have an infinite number of groups. Each value of p $(0 \le p \le 1)$ is a group. Because p is uniform over [0,1], $f(p)=1$. As a result, for a given group of p, the before-event size is one. And for a given group of p, this group's probability to make the event "getting 2 heads out of 3 tosses" happen is a binomial distribution with probability of $C_3^2 p^2 (1-p)$. So the after-event size is

$$\underbrace{\underbrace{k}_{\text{scaling factor}} \ \underbrace{1}_{\substack{\text{before-event} \\ \text{group size}}} \ \underbrace{C_3^2 p^2 (1-p)}_{\substack{\text{the group's probability} \\ \text{to have 2 heads out of 3 tosses}}}}_{\text{After-event size of the groups}}$$

k is a scaling factor such that the sum of the after-event sizes for all the groups is equal to one. Since we have an infinite number of groups, we have to use integration to sum up all the after-event sizes for each group:

$$\int_0^1 k\, C_3^2 p^2 (1-p) dp = 1 \quad \Rightarrow k = \frac{1}{\int_0^1 C_3^2 p^2 (1-p) dp}$$

Then the after-event size (or posterior probability) is:

$$kC_3^2 p^2 (1-p) = \frac{C_3^2 p^2 (1-p)}{\int_0^1 C_3^2 p^2 (1-p) dp} = \frac{p^2 (1-p)}{\int_0^1 p^2 (1-p) dp}$$

It turns out that the posterior probability we just calculated is termed "Beta distribution." Let's generalize this problem:

You are tossing a coin. Not knowing p, the success rate of heads showing up in one toss of the coin, you subjectively assume that p is uniformly distributed over $[0,1]$. Next, you do an experiment by tossing the coin $m+n$ times (where m, n are non-negative integers). You find that, in this experiment, m out of these $m+n$ tosses have heads.

Then the posterior probability of p is:

$$f(p) = \frac{p^m (1-p)^n}{\int_0^1 p^m (1-p)^n dp}$$

The above distribution $f(p)$ is called Beta distribution.

If we set $m = \alpha - 1$ and $n = \beta - 1$ where $\alpha > 0$ and $\beta > 0$, we have

$$f(p)=\frac{p^{\alpha-1}(1-p)^{\beta-1}}{\int_0^1 p^{\alpha-1}(1-p)^{\beta-1}dp}, \text{ where } 0\le p\le 1$$

This is a generalized Beta distribution.

Please note that finding the posterior probability for continuous random variables is not on the Exam P syllabus (it's on Exam C or the new Course 4). However, we introduced it here anyway for two purposes: (1) to get a sense of how to use Bayes' theorem and calculate the posterior probability for a continuous random variable; and more importantly (2) to derive Beta distribution.

Beta distribution is on the Exam P syllabus. Without using the concept of posterior probability, it is very hard for us to intuitively interpret Beta distribution.

We'll pick up Beta distribution later.

Final note. The accuracy of the posterior probability under Bayes' Theorem depends on the accuracy of the prior probability. If the prior probability is way off (because it is based on existing data or purely subjective judgment), the posterior probability will be way off.

Homework for you:

#2, #33 May 2000; #12, #22, 28 Nov 2000; #6, #23 May 2001; #4 Nov 2001; #8, #31 May 2003.

Chapter 8 Random variables

A random variable is a function that assigns a number to each element of the sample space. We typically write a random variable in a capital letter such as X.

Example 1. If we flip a coin and observe which side is up, the sample space is {H, T}. If we assign H=1 and T=0, our random variable is:

$$X = \begin{cases} 1 & \text{with probability of 0.5} \\ 0 & \text{with probability of 0.5} \end{cases}$$

There's more than one way to assign a value to each element in the sample space. In the flip of a coin, we can also assign H=0 and T=1. In this case, the random variable is:

$$X = \begin{cases} 0 & \text{with probability of 0.5} \\ 1 & \text{with probability of 0.5} \end{cases}$$

Of course, you can assign H=2 and T=3, etc.

Example 2. If we flip a coin three times, the sample space is:

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Each element in the sample space has 0.125 chance of occurring. If we assign X =# of heads, Y =# of tails, Z =# of successive flips that have the same outcome, then each element corresponds to the following 3 sets of numerical values:

Element	X	Y	Z	Probability
HHH	3	0	3	0.125
HHT	2	1	2	0.125
HTH	2	1	1	0.125
HTT	1	2	2	0.125
THH	2	1	2	0.125
THT	1	2	1	0.125
TTH	1	2	2	0.125
TTT	0	3	3	0.125

Please note that in HHH, we have 3 consecutive heads; so Z=3. In HTH, no two consecutive outcomes are the same; so Z=1.

Here X, Y, Z are three random variables.

Example 3. If we roll a die and record the side that's face up, the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. We can assign a value to each element in the sample space as follows:

Element of the sample space	X	Probability
1	1	1/6
2	2	1/6
3	3	1/6
4	4	1/6
5	5	1/6
6	6	1/6

Here X is a random variable.

Of course, you can assign values differently as follows:

Element of the sample space	X	Probability
1	6	1/6
2	5	1/6
3	4	1/6
4	3	1/6
5	2	1/6
6	1	1/6

You see that a random variable is really an arbitrary translation of each element in the sample space into a number. Of course, some translation schemes are more useful than others.

What do we gain by such translation? By mapping the entire sample space into a series of numbers, we can extract relevant information from the sample space to solve the problem at hand, while ignoring other details of the sample space.

Example 4. We flip a coin and are interested in finding the # of times heads are up. If we assign 1 and 0 to H and T respectively, then the information about the # of heads up can be conveniently summarized as follows:

$$X = \begin{cases} 1 & \text{with probability of } 0.5 \\ 0 & \text{with probability of } 0.5 \end{cases}$$

You see that we have reduced the coin flipping process to a simple, elegant math equation. More importantly, this equation answers our question at hand.

Example 5. If we flip a coin three times. We are concerned about the # of times heads show up. Let X represent the # of times heads show up, then we:

X	3	2	2	1	2	1	1	0
Probability	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125

Example 6. We roll a die and record the side that's face up. We are interested in finding the probability of getting 1,2,3,4,5, and 6 respectively. If we let random variable X represent the number that's face up, then we have:

X	1	2	3	4	5	6
Probability	1/6	1/6	1/6	1/6	1/6	1/6

Expressed more succinctly:

$$P(X=n)=\frac{1}{6}, \text{ where } n=1,2,3,4,5,6$$

Discrete random variable vs. continuous random variable

If a random variable can take on discrete values, then it's a discrete random variable. If a random variable can take on any value in a range, then it's a continuous random variable.

Example 7. Let the random variable X represent the # of heads we get from flipping a coin n times. Then X can take on integer values ranging from 0 to n. X is a discrete random variable.

Example 8. Let random variable Y represent the number randomly chosen from the range [0,1]. Then Y can take on any value in [0,1]. Y is a continuous random variable.

PMF and CDF for discrete random variables

Probability mass function

The most important way to describe a discrete random variable is through the probability mass function (PMF). If x is a possible value of the random variable X, the probability mass of x, denoted as $p_X(x)$, is the probability that $X=x$:

$$p_X(x)=P(X=x)$$

Example 9. We flip a coin twice and record the # of times we get heads. Let X represent the # of heads in 2 flips of a coin. The probability mass function of X is:

$$p_X(x)=\begin{cases}1/4 & \text{if } x=0\\ 1/2 & \text{if } x=1\\ 1/4 & \text{if } x=2\end{cases}$$

A probability mass function must satisfy the 3 axioms:

- $p_X(x)\geq 0$
- $p_X(x=a\cup x=b)=p_X(a)+p_X(b)$ where $a\neq b$
- $\sum_x p_X(x)=1$

The second condition is trivial. Since $a\neq b$, we have $p_X(x=a\cap x=b)=0$.
So $p_X(x=a\cup x=b)=p_X(a)+p_X(b)$ automatically holds.

So a valid PMF needs to satisfy the following two conditions: $p_X(x)\geq 0$, $\sum_x p_X(x)=1$

Example 10. You are given the following PMF:

$$p_N(n)=e^{-\lambda}\frac{\lambda^n}{n!}, \text{ where } n=0,1,2,...,+\infty \text{ and } \lambda \text{ is a positive constant}$$

Verify that this is a legitimate PMF.

Solution

$$p_N(n)=e^{-\lambda}\frac{\lambda^n}{n!}\geq 0 \text{ for } n=0,1,2,...,+\infty$$

$$\sum_{n=0}^{+\infty}p_N(n)=\sum_{n=0}^{+\infty}e^{-\lambda}\frac{\lambda^n}{n!}=e^{-\lambda}\sum_{n=0}^{+\infty}\frac{\lambda^n}{n!}$$

$$\sum_{n=0}^{+\infty}\frac{\lambda^n}{n!}=1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+...+\frac{\lambda^n}{n!}+...=e^{\lambda} \quad \text{(Taylor series)}$$

$$\Rightarrow \sum_{n=0}^{+\infty}p_N(n)=e^{-\lambda}\sum_{n=0}^{+\infty}\frac{\lambda^n}{n!}=e^{-\lambda}\left(e^{\lambda}\right)=1$$

So $p_N(n)=e^{-\lambda}\dfrac{\lambda^n}{n!}$ is a valid PMF.

Example 11. A special die has 3 sides painted 1, 2, and 3 respectively. If the die is thrown, each side has an equal chance of landing face up on the ground. Two dies are thrown together and let X represent the sum of the two sides facing up.

Find the probability mass function of X.

Solution

	outcome of 2nd throw		
outcome of the 1st throw	1	2	3
1	**2**	**3**	**4**
2	**3**	**4**	**5**
3	**4**	**5**	**6**

In the above table, the blue cells represent the values of X. Because each side has $1/3$ chance of landing face up, each cell has $(1/3)^2=1/9$ chance of occurring.

We convert the above table into the new table below:

x	2	3	4	5	6
$p_X(x)$	$1/9$	$2/9$	$3/9$	$2/9$	$1/9$

To understand the above table, let's look at $p_X(3)=2/9$. This is how we get $p_X(3)=2/9$. There are two ways to have $X=3$: you get a 1 from the 1st die and 2 from the 2nd die (with probability of $1/9$); you get 2 from the 1st die and 1 from the 2nd die (with probability of $1/9$). So the total probability of having $X=3$ is $2/9$.

Example 12. Claim payment, X, has the following PMF:

x	\$0	\$50	\$80	\$135	\$250	\$329
$p_X(x)$	0.32	0.2	0.18	0.1	0.15	0.05

Calculate

1. $P(X>120)$
2. $P(X\le 300 \mid X>120)$

Solution

$$P(X>120)=P(x=135)+P(x=250)+P(x=329)=0.1+0.15+0.05=0.3$$

$$P(X\le 300 \mid X>120)=\frac{P(X>120\cap X\le 300)}{P(X>120)}=\frac{P(120<X\le 300)}{P(X>120)}$$

$$P(120<X\le 300)=P(x=135)+P(x=250)=0.1+0.15=0.25$$

$$\Rightarrow \quad P(X\le 300 \mid X>120)=\frac{P(120<X\le 300)}{P(X>120)}=\frac{0.25}{0.3}=0.833$$

Cumulative probability function (CDF)

The cumulative function is defined as $F_X(x)=P(X\le x)$

Formulas:

$$\boxed{\begin{aligned}&P(a<X\le b)=F(b)-F(a)\\&P(a\le X\le b)=F(b)-F(a)+p(a)\end{aligned}}$$

Proof.

$\{a<X\le b\}=\{X\le b\}-\{X\le a\}$. $\Rightarrow \quad P(a<X\le b)=F(b)-F(a)$

$P(a\le X\le b)=P(a<X\le b)+p(a)=F(b)-F(a)+p(a)$

Example 13. If the PMF for X is:

x	2	3	4	5	6
$p_X(x)$	$1/9$	$2/9$	$3/9$	$2/9$	$1/9$

Then,

$F(-\infty)=P(X\le -\infty)=0$

The minimum value of x is 2; there's no way to have $X\le -\infty$.

$F(0)=P(X\le 0)=0$. There's no way to have $X\le 0$.

$F(1)=P(X\le 1)=0$. There's no way to have $X\le 1$.

$F(2)=P(X\le 2)=p(2)=1/9$

$F(3)=P(X\le 3)=p(2)+p(3)=1/9+2/9=1/3$

$$F(4)=P(X\le 4)=p(2)+p(3)+p(4)=1/9+2/9+3/9=2/3$$
$$F(5)=P(X\le 5)=p(2)+p(3)+p(4)+p(5)=1/9+2/9+3/9+2/9=8/9$$
$$F(6)=P(X\le 6)=p(2)+p(3)+p(4)+p(5)+p(6)=1/9+2/9+3/9+2/9+1/9=1$$

$$\begin{aligned}F(7)=P(X\le 7)&=p(2)+p(3)+p(4)+p(5)+p(6)+p(7)\\&=1/9+2/9+3/9+2/9+1/9+0=1 \quad \text{because } p(7)=0\end{aligned}$$

$$F(+\infty)=P(X\le +\infty)=P(X\le 6)=1$$

PDF and CDF for continuous random variables

For a continuous random variable X, the probability density function (PDF), $f(x)$, is defined as:

$$P(a\le x\le b)=\int_a^b f(x)dx$$

$P(a\le x\le b)$ is the area under the graph $f(x)$. Because including or excluding the end points doesn't affect the area, including or excluding the end points doesn't affect the probability:

$$P(a<X<b)=P(a\le X<b)=P(a\le X\le b)=P(a<X\le b)=\int_a^b f(x)dx$$

The CDF (cumulative probability function) of the continuous random variable X is defined as:

$F(x)=P(X\le x)$. This is the same definition when X is discrete.

If a random variable is <u>discrete</u>, we say <u>PMF</u> (probability mass function); if a random variable is <u>continuous</u>, we say <u>PDF</u> (probability density function).

Whether a random variable is <u>discrete or continuous</u>, we always say <u>CDF</u> (cumulative probability function).

Please note that often for the sake of convenience, people use $f(x)$ to refer to either PMF $p_X(x)$ or PDF $f(x)$.

Properties of CDF

Rule 1 $F(x)=P(X\le x)$ for all x -- this is just the definition.

Rule 2 CDF can never be decreasing. If $a\le b$, then $F(a)\le F(b)$.

To see why, notice $F(x)=P(X\le x)$. If $a\le b$, then $\{x\le a\}\in\{x\le b\}$. In other words, $x\le b$ contains $x\le a$. So $P(x\le b)\ge P(x\le a)$. This gives us $F(b)\ge F(a)$.

Rule 3 $F(-\infty)=0$ and $F(+\infty)=1$.

They are true for both discrete and continuous random variables. To see why, notice $-\infty<X<+\infty$. There's zero chance that X can be smaller or equal to $-\infty$; $F(-\infty)=P(X\le-\infty)=0$. On the other hand, we are 100% certain that X cannot exceed $+\infty$. So $F(+\infty)=P(X\le+\infty)=1$.

Rule 4 If X is discrete and takes integer values, the PMF and CDF can be obtained from each other by summing or differencing:

$$F(k)=\sum_{i=-\infty}^{k}p_X(i) \text{ -- this is the definition of } F(k)$$

$$p_X(k)=P(X\le k)-P(X\le k-1)=F(k)-F(k-1)$$

Rule 5 If X is continuous, the PDF and CDF can be obtained from each other by integration or differentiation:

$$F(x)=\int_{-\infty}^{x}f(t)dt,\quad f(x)=\frac{d}{dx}F(x).$$

By definition, $F(x)=P(X\le x)=P(-\infty\le X\le x)=\int_{-\infty}^{x}f(t)dt$. Taking the derivative at both sides of $F(x)=\int_{-\infty}^{x}f(t)dt$ gives us $f(x)=\frac{d}{dx}F(x)$.

Example 14. X has the following density: $f(x)=3x^2$ where $0\le x\le 1$.

Then, $F(x)=P(X\le x)=\int_{-\infty}^{x}f(t)dt=\int_{0}^{x}f(t)dt=\int_{0}^{x}3t^2dt=x^3$.

$P(0.2\le X\le 0.6)=F(0.6)-F(0.2)=0.6^3-0.2^3=0.208$

Example 15. A real number is randomly chosen from [0,1]. Then this number is squared. Let X represent the result.

Find the PDF and CDF for X .

Solution

We'll find the CDF first. Let U represent the # randomly drawn from [0,1]. Then $X = U^2$.

$$F_X(x) = P(X \le x) = P(U^2 \le x) = P\left(U \le \sqrt{x}\right)$$

Because any number in the interval [0,1] has an equal chance of being drawn, $P\left(U \le \sqrt{x}\right)$ must be proportional to the length of the interval $\left[0, \sqrt{x}\right]$. The total probability that $P(U \le 1) = 1$ -- we are 100% certain that any number taken from [0,1] must not exceed 1. Consequently,

$$P\left(U \le \sqrt{x}\right) = \text{length of } \left[0, \sqrt{x}\right] = \sqrt{x}$$

$$\Rightarrow \quad F_X(x) = \sqrt{x} \quad \text{where } 0 \le x \le 1$$

If $x \le 0$, then $F_X(x) = 0$; if $x \ge 1$, then $F_X(x) = 1$.
So the CDF is:

$$F_X(x) = \begin{cases} 0 & \text{if } x \le 0 \\ \sqrt{x} & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x \ge 1 \end{cases}$$

$$\Rightarrow \quad f(x) = \frac{d}{dx} F(x) = \frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}} \quad \text{where } 0 < x \le 1 \text{ and } f(x) \text{ is zero elsewhere.}$$

Please note that the following key difference between PMF for a discrete random variable and PDF for a continuous random variable:

PMF is a real probability and its value must not exceed one; PDF is a fake probability and can take on any non-negative value. PDF itself doesn't have any meaning. For PDF to be useful, we must integrate it over a range.

In the example above, PDF is $f(x)=\frac{1}{2\sqrt{x}}$ for $0<x\le 1$. When $x\to 0$, $f(x)=\frac{1}{2\sqrt{x}}\to +\infty$. $f(x)=\frac{1}{2\sqrt{x}}$ is not a probability. To get a probability, we must integrate $f(x)=\frac{1}{2\sqrt{x}}$ over a range. For example, if we integrate $f(x)$ over $[a,b]$, we'll get a real probability:

$$P(a<X\le b)=\int_a^b f(x)dx$$

Mean and variance of a random variable

You just have to memorize a series of formulas:

If X is discrete, then

mean $E(X)=\sum_x x\, p_X(x)$

variance $Var(X)=E[X-E(X)]^2=\sum_x [x-E(X)]^2 p_X(x)=E(X^2)-E^2(X)$

If X is continuous, then the

mean $E(X)=\int_{-\infty}^{+\infty} xf(x)dx$

variance $Var(X)=E[X-E(X)]^2=\int_{-\infty}^{+\infty}[x-E(X)]^2 f(x)dx=E(X^2)-E^2(X)$

Standard deviation of X - no matter X is continuous or discrete

$\sigma_X=\sqrt{Var(X)}$

Example 16.

x	1	2	3	4	5
$p(x)$	0.2	0.15	0.05	0.43	0.17

Then

$$E(X)=\sum_x x\, p_X(x)=1(0.2)+2(0.15)+3(0.05)+4(0.43)+5(0.17)=3.22$$

$$Var(X)=E\left[X-E(X)\right]^2$$
$$=0.2(1-3.22)^2+0.15(2-3.22)^2+0.05(3-3.22)^2+0.43(4-3.22)^2+0.17(5-3.22)^2$$
$$=2.0116$$

$$\sigma_X=\sqrt{Var(X)}=1.4183$$

Alternative way to calculate $Var(X)$:

$$E(X^2)=\sum_x x^2 p_X(x)=1^2(0.2)+2^2(0.15)+3^2(0.05)+4^2(0.43)+5^2(0.17)=12.38$$
$$Var(X)=E(X^2)-E^2(X)=12.38-3.22^2=2.0116$$

$$\sigma_X=\sqrt{Var(X)}=1.4183$$

Example 17. $f(x)=3x^2$ where $0\le x\le 1$. Then

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_0^1 x(3x^2)dx=\frac{3}{4}$$
$$Var(X)=E\left[X-E(X)\right]^2=\int_{-\infty}^{+\infty}\left[x-E(X)\right]^2 f(x)=\int_0^1 (x-3/4)^2 3x^2dx=\frac{3}{80}$$
$$\sigma_X=\sqrt{Var(X)}=\sqrt{\frac{3}{80}}$$

Alternative way to calculate $Var(X)$:

$$E\left(X^2\right)=\int_{-\infty}^{+\infty} x^2 f(x)dx=\int_0^1 x^2\left(3x^2\right)dx=\frac{3}{5}$$

$$Var(X)=E\left(X^2\right)-E^2(X)=\frac{3}{5}-\left(\frac{3}{4}\right)^2=\frac{3}{80}$$

Mean of a function

Many times we need to find $E\left[Y=g(X)\right]$. One way to find $E[Y]$ is to find the pdf $f_Y(y)=f_Y\left[g(x)\right]$ and then calculate $E(Y)=\int_{-\infty}^{+\infty} y\ f_Y(y)dy$. However, finding $f_Y(y)=f_Y\left[g(x)\right]$ is not easy. Fortunately, we can calculate $E[Y]$ without finding $f_Y(y)=f_Y\left[g(x)\right]$:

$$E\left[Y=g(X)\right]=\sum_x g(x)\,p_X(x) \quad \text{if } X \text{ is discrete}$$

$$E\left[Y=g(X)\right]=\int_{-\infty}^{+\infty} g(x)\,f(x)dx \quad \text{if } X \text{ is continuous}$$

Don't worry about how to prove it. Just memorize it.

Example 18. $Y=X^2-1$, where X has the following distribution:

$f(x)=e^{-x}$, where $x>0$.

Then $E\left(X^2-1\right)=\int_0^{+\infty}\left(x^2-1\right)e^{-x}dx=\int_0^{+\infty} x^2e^{-x}dx-\int_0^{+\infty} e^{-x}dx$

$$\int_0^{+\infty} x^2e^{-x}dx=-\int_0^{+\infty} x^2de^{-x}=-x^2e^{-x}\Big|_0^{+\infty}+\int_0^{+\infty} e^{-x}dx^2=\int_0^{+\infty} e^{-x}dx^2=2\int_0^{+\infty} e^{-x}xdx$$

$$\int_0^{+\infty} e^{-x}xdx=-\int_0^{+\infty} xde^{-x}=-xe^{-x}\Big|_0^{+\infty}+\int_0^{+\infty} e^{-x}dx=\int_0^{+\infty} e^{-x}dx=-e^{-x}\Big|_0^{+\infty}=1$$

$$\Rightarrow\quad E\left(X^2-1\right)=\int_0^{+\infty}\left(x^2-1\right)e^{-x}dx=1$$

Alternative method:

e^{-x} is the exponential pdf (probability density function) with mean $\theta = 1$. Consequently,

$$\int_0^{+\infty} x^2 e^{-x} dx = E\left(X^2\right) = E^2\left(X\right) + Var\left(X\right) = \theta^2 + \theta^2 = 2\theta^2 = 2$$

$$E\left(X^2 - 1\right) = E\left(X^2\right) - E\left(1\right) = 2 - 1 = 1$$

Properties of mean:

$$E\left(aX + b\right) = a\, E\left(x\right) + b$$
$$E\left(X + Y\right) = E\left(X\right) + E\left(Y\right)$$

$E\left(XY\right) = E\left(X\right)E\left(Y\right)$ -- if X and Y are independent.

Example 19. X has the following distribution:

X	1	2	3	4
$p_X\left(x\right)$	0.15	0.2	0.3	0.35

$Y = X^2 - 1$. Calculate $Var\left(Y\right)$.

Solution

Y^2	0	9	64	225
$Y = X^2 - 1$	0	3	8	15
X	1	2	3	4
$p_X\left(x\right)$	0.15	0.2	0.3	0.35

$$E\left[Y = g\left(X\right)\right] = \sum_x g\left(x\right) p_X\left(x\right) = 0\left(0.15\right) + 3\left(0.2\right) + 8\left(0.3\right) + 15\left(0.35\right) = 8.25$$

$$E\left[Y^2\right] = \sum_x g^2\left(x\right) p_X\left(x\right) = 0\left(0.15\right) + 9\left(0.2\right) + 64\left(0.3\right) + 225\left(0.35\right) = 99.75$$

$$Var\left(Y\right) = E\left(Y^2\right) - E^2\left(Y\right) = 99.75 - 8.25^2 = 31.6875$$

Chapter 9 Independence

Event A and Event B are independent if and only if

$$P(A \cap B) = P(A) \times P(B)$$

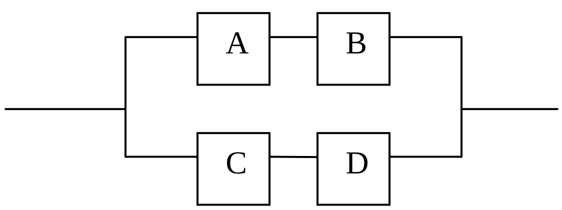
Events $A_{1,}\ A_{2,}\ ...,\ A_n$ are mutually independent if and only if for any $2 \leq k \leq n$ and for any subset of indices $i_1, i_2, ... i_k$,

$$P\left(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}\right) = P\left(A_{i_1}\right) P\left(A_{i_2}\right) ... P\left(A_{i_k}\right)$$

Sample Problems and Solutions

Problem 1

A system consists of four independent components connected as in the diagram below. The system works as long as electric current can flow through the system from left to right. Each component has a failure rate of 5%, independent of any other components. Determine the probability that the system fails.



Solution

Let's first calculate the probability that the system works.

Let the event A represent that the component A works,
Let the event B represent that the component B works,
Let the event C represent that the component C works,
Let the event D represent that the component D works.

The event that the system works is $S = (A \cap B) \cup (C \cap D)$.

Use the formula $P(M \cup N) = P(M) + P(N) - P(M \cap N)$:

$$P(S)=P\left[(A\cap B)\cup(C\cap D)\right]=P(A\cap B)+P(C\cap D)-P\left[(A\cap B)\cap(C\cap D)\right]$$

$P(A\cap B)=P(A)P(B)$ (because A,B are independent)
$P(C\cap D)=P(C)P(D)$ (because C,D are independent)
$P\left[(A\cap B)\cap(C\cap D)\right]=P(A\cap B)P(C\cap D)$
(because $(A\cap B),(C\cap D)$ are independent)

$$P(S)=P(A\cap B)+P(C\cap D)-P\left[(A\cap B)\cap(C\cap D)\right]$$
$$=P(A)P(B)+P(C)P(D)-P(A)P(B)P(C)P(D)$$
=0.95*0.95+0.95*0.95-0.95*0.95*0.95*0.95
=0.99

The probability that the systems fails is

$P\left(\overline{S}\right)=1-P(S)$=1-0.99=0.01

Problem 2

Two random variables X,Y have zero correlation (i.e. $\rho_{X,Y}=0$). Are X,Y independent?

Solution

If X,Y are independent, then $\rho_{X,Y}=0$. However, the reverse may not be true; zero correlation doesn't automatically mean independence.

Correlation coefficient is defined as

$$\rho_{X,Y}=\frac{\operatorname{cov}(X,Y)}{\sigma_X\,\sigma_Y}$$

$\rho_{X,Y}$ measures whether X,Y have a good linear relationship. $\rho_{X,Y}=1$ or $\rho_{X,Y}=-1$ means that X,Y have a perfect linear relationship of $Y=aX+b$ or $X=cY+d$ (where a,b,c,d are constants).

$\rho_{X,Y}=0$ indicates that X,Y don't have any linear relationship. However, X,Y may have a non-linear relationship (such as $Y=X^2$). As a result, even $\rho_{X,Y}=0$, X,Y may not be independent from each other.

One exception to the rule (see Chapter 25 on bivariate normal distribution): when X, Y are both normal random variables, then $\rho_{X,Y} = 0$ means that X, Y to are indeed independent.

Problem 3

Two events A, B have no intersections. In other words, $A \cap B = \varnothing$. Does this mean that A, B are independent?

Solution

$A \cap B = \varnothing$ doesn't mean that A, B are independent.

$A \cap B = \varnothing$ merely means that A, B are mutually exclusive. Mutually exclusive events may be affected by a common factor, in which case A, B are not independent.

For example, let A represent that an exam candidate passes Exam P on the first try; let B represent that the same candidate passes Exam P after at least a second try. We can see that $A \cap B = \varnothing$, but A, B are not independent. If A occurs, B definitely cannot occur. Likewise, if B occurs, A definitely cannot occur.

If A, B are truly independent, then whether A occurs or not does not have any influence on whether B will occur. In other words, if A, B are independent, then the information that A occurs is useless in predicting whether B may occur or not; the information that B occurs is useless in predicting whether A may occur or not.

The only way to test whether A, B are independent is to check whether the condition $P(A \cap B) = P(A)P(B)$ holds. If $P(A \cap B) = P(A)P(B)$, then A, B are independent; if $P(A \cap B) \neq P(A)P(B)$, then A, B are not independent.

If $A \cap B = \varnothing$, then $P(A \cap B) = 0$. $P(A \cap B) = P(A)P(B)$ holds only if either $P(A) = 0$ or $P(B) = 0$ or both. We can see that if $A \cap B = \varnothing$, then the only way for A, B to be independent is that at least one of the two events A, B is empty. This conclusion is common sense. An event doomed never to occur (independent of any other events occurring or not) is indeed an independent event.

Problem 4

If $P(A \cap B \cap C) = P(A)P(B)P(C)$, does this mean that A, B, C are independent?

Solution

$P(A \cap B \cap C) = P(A)P(B)P(C)$ does not mean that A, B, C are independent.

A, B, C are independent if the following conditions are met:

$$P(A \cap B) = P(A)P(B)$$
$$P(A \cap C) = P(A)P(C)$$
$$P(B \cap C) = P(B)P(C)$$
$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

Homework for you – redo all the problems listed in this chapter.

Chapter 10 Percentile, mean, median, mode, moment

MEAN of $X = E(X)$.

Percentile. If your SAT score in the 90-th percentile, then 90% of the people who took the same test scored below you or got the same score as you did; only 10% of the people scored better than you.

For a random variable X, p-th percentile (donated as x_p), means that

$$\Pr(X \leq x_p) = p\% \quad \Leftrightarrow F(x_p) = p\% \quad \Leftrightarrow \Pr(X > x_p) = 1 - p\%$$

Median=Middle=50$^{\text{th}}$ percentile $= x_{50}$

MOde=Most Often=Most Observed $\Rightarrow f(x)$ is maxed at mode $\Rightarrow \left.\dfrac{d}{dx} f(x)\right|_{\text{Mode}} = 0$

k-th Moment of $X = E(X^k)$. 1$^{\text{st}}$ moment$=E(X)$. 2$^{\text{nd}}$ moment$=E(X^2)$

k-th Central Moment of $X = E\left[X - E(X)\right]^k$. 1$^{\text{st}}$ central moment=0. 2$^{\text{nd}}$ central moment$=Var(X)$.

Sample Problems and Solutions

Problem 1 (SOA #22 May 2003)

An insurer's annual weather related-loss, X, is a random variable with density function

$$f(x) = \begin{cases} \dfrac{2.5(200)^{2.5}}{x^{3.5}} & \text{for } x > 200 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the difference between the 30$^{\text{th}}$ percentile and 70$^{\text{th}}$ percentile.

Solution

Let x_{30} and x_{70} represent the 30th and 70th percentile.

$$F(x)=\int_{200}^{x} f(t)dt=1-\left(\frac{200}{x}\right)^{2.5}$$

$$F\left(x_{30}\right)=0.3 \quad \Rightarrow \quad 1-\left(\frac{200}{x_{30}}\right)^{2.5}=0.3, \quad x_{30}=230.67$$

$$F\left(x_{70}\right)=0.7 \quad \Rightarrow 1-\left(\frac{200}{x_{70}}\right)^{2.5}=0.7 \quad \Rightarrow \left(\frac{200}{x_{70}}\right)^{2.5}=0.3 \quad \Rightarrow x_{70}=323.73$$

The difference is $323.73-230.67=93.06$

Problem 2 (#18 May 2000)

An insurance company policy reimburses dental expenses, X, up to a maximum benefit of 250. The probability density function for X is:

$$f(x)=\begin{cases} c\,e^{-0.004\,x} & \text{for } x\geq 0 \\ 0 & \end{cases}$$

where c is a constant. Calculate the median benefit for this policy.

(A) 161, (B) 165, (C) 173 (D) 182 (E) 250.

Solution

First, we need to create a random variable representing benefits payable under the policy.

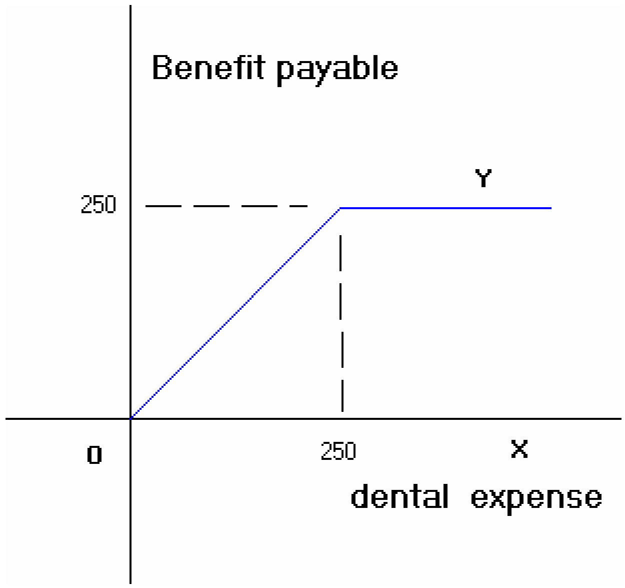
Let Y = benefits payable under the policy.

$$Y=\begin{cases} X & \text{If } X\leq 250 \\ 250 & \text{If } X>250 \end{cases}$$

Median=Middle=50th percentile = y_{50}

$\Pr(Y \le y_{50}) = 50\% \quad \Leftrightarrow F(y_{50}) = 50\%$. We need to find y_{50}.

It is always a good idea to draw a diagram.



We can see that Y is a non-decreasing function of X. A non-decreasing transformation preserves the percentiles. In other words, if x_p is the p-th percentile of X and y_p is the p-th percentile of Y, then $y_p = y\left(x_p\right)$

$\Pr(Y \le y_{50}) = \Pr(X \le x_{50}) = 50\%$ (i.e. x_{50} corresponds to y_{50})
So we just need to find x_{50}, the 50th percentile of X. Then $Y\left(x_{50}\right)$ is the 50th percentile of Y.

Please note that X has exponential distribution.
So $F(x) = 1 - e^{-0.004\,x}$.

$$F(x_{50}) = 50\% \quad \Rightarrow 1 - e^{-0.004\,x_{50}} = 50\% \quad \Rightarrow e^{-0.004\,x_{50}} = 50\%$$

$$-0.004\,x_{50} = \ln\left(50\%\right) \quad \Rightarrow x_{50} = -\frac{\ln\left(50\%\right)}{0.004} = 173.28$$

$y(x = 173.28) = 173.28 \quad$ (because $173.28 < 250$)
So y_{50} =173.28.

Final point. $F(250)=1-e^{-0.004\times 250}=0.632$. So 250 is the 63.2-th percentil of the dental expense X; 250 is also the 63.2 percentile of Y. Because Y is a always \$250 once $X \geq 250$, any percentile higher than 63.2 (such as 70 percentile of Y, 75 percentile of Y, 99 percentile of Y) is always 250.

Problem 3 (#39 Course 2 Sample Test)

The loss amount, X, for a medical insurance policy has cdf

$$F(x)=\begin{cases} 0 & x<0 \\ \dfrac{1}{9}\left(2x^2-\dfrac{1}{3}x^3\right) & 0\leq x\leq 3 \\ 1 & x>3 \end{cases}$$

Calculate the mode of the distribution.

(A) 2/3, (B) 1, (C) 3/2 (D) 2 (E) 3.

Solution

MOde=Most Observed $\Rightarrow f(x)$ is maxed at mode $\Rightarrow \dfrac{d}{dx}f(x)\Big|_{\text{Mode}}=0$

$f(x)=\dfrac{d}{dx}F(x)=\dfrac{1}{9}\left(4x-x^2\right)$ for $0\leq x\leq 3$ (anywhere else =0)

$$\frac{d}{dx}f(x)=\frac{d}{dx}\frac{1}{9}\left(4x-x^2\right)=\frac{1}{9}\left(4-2x\right)$$

$\dfrac{d}{dx}f(x)$ is zero at $x=2$. So the mode=2.

Homework for you – redo all the problems listed in this chapter.

Chapter 11 Find $E(X), Var(X), E(X|Y), Var(X|Y)$

Calculating $E(X), Var(X), E(X|Y), Var(X|Y)$ for a discrete random variable is a commonly tested type of problem. Developing the skill to calculate $E(X)$ and $Var(X)$ quickly and accurately will help you on many SOA/CAS exams.

Because the BA II Plus Statistic Worksheet can calculate $E(X)$ and $Var(X)$ of a discrete random variable X when X has no more than 50 distinct values (most exam problems fall within this limit), there is no need to calculate the mean and variance using the standard formula under exam conditions.

The mean and variance may also be calculated using the TI-30 IIS and the formula:

$$E(X) = \sum x f(x), \quad E(X^2) = \sum x^2 f(x), \quad Var(X) = E(X^2) - E^2(X)$$

Sample Problems and Solutions

Problem 1 (#8 May 2000) A probability distribution of the claim sizes for an auto insurance policy is given in the table below:

Claim Size	Probability
20	0.15
30	0.10
40	0.05
50	0.20
60	0.10
70	0.10
80	0.30

What percentage of the claims is within one standard deviation of the mean claim size?

(A) 45%, (B) 55%, (C) 68%, (D) 85%, (E)100%

Solution

This problem is conceptually easy but calculation-intensive. It is easy to make calculation errors. Always let the calculator do all the calculations for you.
Using BA II Plus

One critical thing to remember about the BA II Plus Statistics Worksheet is that you cannot directly enter the probability mass function $f(x_i)$ into the calculator to find $E(X)$ and $Var(X)$. BA II Plus 1-V Statistics Worksheet accepts only scaled-up probabilities that are positive integers. If you enter a non-integer value to the statistics worksheet, you will get an error when attempting to retrieve $\mathrm{E}(\mathrm{X})$ and $\mathrm{Var}(\mathrm{X})$.

To overcome this constraint, first scale up $f(x_i)$ to an integer by multiplying $f(x_i)$ by a common integer.

Claim Size x	Probability $\Pr(x)$	Scaled-up probability $=100\Pr(x)$
20	0.15	15
30	0.10	10
40	0.05	5
50	0.20	20
60	0.10	10
70	0.10	10
80	0.30	30
Total	1.00	100

Next, enter the 7 data pairs of (claim size and scaled-up probability) into the BA II Plus Statistics Worksheet to get $E(X)$ and σ_X .

BA II Plus calculator key sequences:

Procedure	Keystrokes	Display
Set the calculator to display 4 decimal places	2nd [FORMAT] 4 ENTER	DEC=4.0000
Set AOS (Algebraic operating system)	2nd [FORMAT], keep pressing ↓ multiple times until you see "Chn." Press 2nd [ENTER] (if you see "AOS", your calculator is already in AOS, in which case press [CLR Work])	AOS
Select data entry portion of Statistics worksheet	2nd [Data]	X01 (old contents)
Clear worksheet	2nd [CLR Work]	X01 0.0000
Enter data set	20 ENTER ↓ 15 ENTER	X01=20.0000 Y01=15.0000
	↓ 30 ENTER ↓ 10 ENTER	X02=30.0000 Y02=10.0000

	↓ 40 ENTER ↓ 5 ENTER	X03=40.0000 Y03=5.0000
	↓ 50 ENTER ↓ 20 ENTER	X04=50.0000 Y04=20.0000
	↓ 60 ENTER ↓ 10 ENTER	X05=60.0000 Y05=10.0000
	↓ 70 ENTER ↓ 10 ENTER	X06=70.0000 Y06=10.0000
	↓ 80 ENTER ↓ 30 ENTER	X07=80.0000 Y07=30.0000
Select statistical calculation portion of Statistics worksheet	2nd [Stat]	Old content
Select one-variable calculation method	Keep pressing 2nd SET until you see 1-V	1-V
View the sum of the scaled-up probabilities	↓	n=100.0000 (Make sure the sum of the scaled-up probabilities is equal to the scaled-up common factor, which in this problem is 100. If n is not equal to the common factor, you've made a data entry error.)
View mean	↓	$\bar{x}$ =55.0000
View sample standard deviation	↓	S_x =21.9043 (this is a sample standard deviation--- don't use this value). Note that $S_x = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
View standard deviation	↓	σ_X =21.7945
View $\sum X$	↓	$\sum X$ =5,500.0000 (not needed for this problem)
View $\sum X^2$	↓	$\sum X^2$ =350,000.0000 (not needed for this problem, though this function might be useful for other calculations)

You should always double check (using ↑↓ to scroll up or down the data pairs of X and Y) that your data entry is correct before accepting $E(X)$ and σ_X generated by BA II Plus.

If you have made an error in data entry, you can 2nd DEL to delete a data pair (X, Y) or 2nd INS to insert a data pair (X,Y). If you typed a wrong number, you can use → to delete the wrong number and then re-enter the correct number. Refer to the BA II Plus guidebook for details on how to correct data entry errors.

If this procedure of calculating $E(X)$ and σ_X seems more time-consuming than the formula-driven approach, it could be because you are not familiar with the BA II Plus Statistics Worksheet yet. With practice, you will find that using the calculator is quicker than manually calculating with formulas.

Then, we have

$$(\mu_X - \sigma_X, \mu_X + \sigma_X) = (55 - 21.7945, \quad 55 + 21.7945)$$
$$=(33.21, \ 76.79)$$

Finally, you find

$$\Pr(33.21 \le X \le 76.79) = \Pr(X = 40) + \Pr(X = 50) + \Pr(X = 60) + \Pr(X = 70)$$
$$=0.05+0.20+0.10+0.10 = 0.45$$

Using TI-30X IIS

Though the TI-30X IIS statistic function can also solve this problem, we will use a standard formula-driven approach. This generic approach takes advantage of the TI-30X IIS "redo calculation" functionality.

First, calculate $E(X)$ using $E(X) = \sum xf(x)$. Then modify the formula

$\sum xf(x)$ to $\sum x^2 f(x)$ to calculate Var(X) without re-entering $f(x)$.

To find $E(X)$, we type:

20*.15+30*.1+40*.05+50*.2+60*.1+70*.1+80*.3

Then press "Enter." $E(X)$=55.

Next we modify the formula
$20*.15+30*.1+40*.05+50*.2+60*.1+70*.1+80*.3$

to

$20^2*.15+30^2*.1+40^2*.05+50^2*.2+60^2*.1+70^2*.1+80^2*.3$

To change 20 to 20^2, move the cursor immediately to the right of the number "20" so your cursor is blinking on top of the multiplication sign $*$. Press "2nd" "INS" "x^2".

You find that

$20^2*.15+30^2*.1+40^2*.05+50^2*.2+60^2*.1+70^2*.1+80^2*.3$
=3500

So $E(X^2)$ =3,500

$Var(X)=E(X^2)-E^2(X)$=3,500-55^2=475.

Finally, you can calculate σ_X and the range of $(\mu_X-\sigma_X, \mu_X+\sigma_X)$.

Keep in mind that you can enter up to 88 digits for a formula in TI-30X IIS. If your formula exceeds 88 digits, TI 30X IIS will ignore the digits entered after the 88th digit.

Problem 2 (#19, November 2001)

A baseball team has scheduled its opening game for April 1. If it rains on April 1, the game is postponed and will be played on the next day that it does not rain. The team purchases insurance against rain. The policy will pay 1,000 for each day, up to 2 days, that the opening game is postponed. The insurance company determines that the number of consecutive days of rain beginning on April 1 is a Poisson random variable with a 0.6 mean. What is the standard deviation of the amount the insurance company will have to pay?

(A) 668, (B) 699, (C) 775, (D) 817, (E) 904

Solution

Let N =# of days it rains consecutively. N can be 0,1,2, …or any non-negative integer.

$$\Pr(N=n)=e^{-\lambda}\frac{\lambda^n}{n!}=e^{-0.6}\frac{0.6^n}{n!} \quad (n=0,1,2,..+\infty)$$

If you don't understand this formula, don't worry; you'll learn it later. For now, take the formula as given and focus on the calculation.

Let X = payment by the insurance company. According to the insurance contract, if there is no rain (n=0), X=0. If it rains for only 1 day, X=\$1,000. If it rains for two or more days in a row, X is always \$2,000. We are asked to calculate σ_X .

If a problem asks you to calculate the mean, standard deviation, or other statistics of a discrete random variable, it is always a good idea to list the variables' values and their corresponding probabilities in a table before doing the calculation to organize your data. So let's list the data pair (X , probability) in a table:

Payment X	Probability of receiving X
0	$\Pr\left(N=0\right)=e^{-0.6}\dfrac{0.6^0}{0!}=e^{-0.6}$
1,000	$\Pr\left(N=1\right)=e^{-0.6}\dfrac{0.6^1}{1!}=0.6e^{-0.6}$
2,000	$\Pr(N\geq 2)=\Pr(N=2)+\Pr(N=3)+...$ $=1-[\Pr(N=0)+\Pr(N=1)]$ $=1-1.6e^{-0.6}$

Once you set up the table above, you can use BA II Plus's Statistics Worksheet or TI-30 IIS to find the mean and variance.

<u>Calculation Method 1 --- Using TI-30X IIS</u>

First we calculate the mean by typing:

1000*.6e^(-.6)+2000(1-1.6e^(-.6

As mentioned before, when typing e^(-.6) for $e^{-0.6}$, you need to use the negative sign, not the minus sign, to get "-6." If you type the minus sign in e^($-$.6), you will get an error message.

Additionally, for $0.6e^{-0.6}$, you do not need to type 0.6*e^(-.6), just type .6e^(-.6). Also, to calculate $2000(1-1.6e^{-.6})$, you do not need to type 2000*(1-1.6*(e^(-.6))). Simply type

2000(1-1.6e^(-.6

Your calculator understands you are trying to calculate $2000(1-1.6e^{-.6})$. However, the omission of the parenthesis sign works only for the last item in your formula. In other words, if your equation is

$$2000(1-1.6e^{-.6})+1000\times.6e^{-.6}$$

you have to type the first item in its full parenthesis, but can skip typing the closing parenthesis in the 2nd item:

2000(1-1.6e^(-.6)) + 1000*.6e^(-.6

If you type

2000(1-1.6e^(-.6 + 1000*.6e^(-.6

your calculator will interpret this as

2000(1-1.6e^(-.6 + 1000*.6e^(-.6)))

Of course, this is not your intention.

Let's come back to the calculation. After you type

1000*.6e^(-.6)+2000(1-1.6e^(-.6

press "ENTER." You should get $E(X)=573.0897$. This is an intermediate value. You can store it on your scrap paper or in one of your calculator's memories.

Next, modify your formula to get $E(x^2)$ by typing:

$$1000^2*.6e\wedge(-.6)+2000^2*(1-1.6\wedge(-.6$$

You will get 816892.5107. This is $E(x^2)$. Next, calculate $Var(X)$

$$Var(X)=E(x^2)-E^2(x)=488460.6535$$
$$\sigma_X=\sqrt{Var(x)}=698.9960.$$

<u>Calculation Method 2 --Using BA II Plus</u>

First, please note that you can always calculate σ_X without using the BA II Plus built-in Statistics Worksheet. You can calculate $E(X), E(X^2), Var(X)$ in BA II Plus as you do any other calculations without using the built-in worksheet.

In this problem, the equations used to calculate σ_X are:

$$E(X)=0*e^{-.6}+1,000(.6e^{-.6})+2,000(1-1.6e^{-.6})$$

$$E(X^2) = 0^2 \times e^{-.6} + 1,000^2 \times .6e^{-.6} + 2,000^2(1-1.6e^{-.6})$$

$$Var(x) = E(x^2) - E^2(x),\ \ \sigma_X = \sqrt{Var(x)}$$

You simply calculate each item in the above equations with BA II Plus. This will give you the required standard deviation.

However, we do not want to do this hard-core calculation in an exam. BA II Plus already has a built-in statistics worksheet and we should utilize it.
The key to using the BA II Plus Statistics Worksheet is to scale up the probabilities to integers. To scale the three probabilities:

$$\left(e^{-.6},\ \ 0.6e^{-.6},\ \ 1-1.6e^{-.6}\right)$$

is a bit challenging, but there is a way:

Payment X	Probability (assuming you set your BA II Plus to display 4 decimal places)	Scale up probability to integer (multiply the original probability by 10,000)
0	$e^{-0.6}=$ 0.5488	5,488
1,000	$0.6e^{-0.6}=$ 0.3293	3,293
2,000	$1-1.6e^{-0.6}=$0.1219	1,219
Total	1.0	10,000

Then we just enter the following data pairs into BA II Plus's statistics worksheet:

X01=0 Y01=5,488;
X02=1,000 Y02=3,293;
X03=2,000 Y03=1,219.

Then the calculator will give you $\sigma_X = 698.8966$

Make sure your calculator gives you *n* that matches the sum of the scaled-up probabilities. In this problem, the sum of your scaled-up probabilities is 10,000, so you should get *n*=10,000. If your calculator gives you *n* that is not 10,000, you know that at least one of the scaled-up probabilities is wrong.

Of course, you can scale up the probabilities with better precision (more closely resembling the original probabilities). For example, you can scale them up this way (assuming you set your calculator to display 8 decimal places):

Payment X	Probability	Scale up probability to integer more precisely (multiply the original probability by 100,000,000)
0	$e^{-0.6}=$ 0.54881164	54,881,164
1,000	$0.6e^{-0.6}=$ 0.32928698	32,928,698
2,000	$1\text{-}1.6e^{-0.6}$ =0.12190138	12,190,138
Total		100,000,000

Then we just enter the following data pairs into BA II Plus's statistics worksheet:

X01=0 Y01=54,881,164;
X02=1,000 Y02=32,928,698;
X03=2,000 Y03=12,190,138.

Then the calculator will give you σ_X =698.8995993 (remember to check that *n*=100,000,000)

For exam problems, scaling up the original probabilities by multiplying them by 10,000 is good enough to give you the correct answer. Under exam conditions it is unnecessary to scale the probability up by multiplying by100,000,000.

By now the shortcomings of the TI-30X IIS statistics function should be apparent. TI-30 IIS has a limit that the scaled-up probability must not
exceed 99. If you use the TI-30X IIS statistics function to solve the problem, this is your table for scaled-up probabilities:

Payment X	Probability (assuming TI-30X IIS displays 2 decimals)	Scale up probability to integer (multiply the original probability by 100. If you multiply probabilities with 1000, you won't be able to enter the scaled-up probabilities on TI-30X IIS)
0	$e^{-0.6}=$ 0.55	55
1,000	$0.6e^{-0.6}=$ 0.33	33
2,000	$1\text{-}1.6e^{-0.6}$ =0.12	12
Total	1.00	100

If you enter the 3 data pairs (X1=0,FRQ=55), (X2=1000,FRQ=33), (X3=2000,FRQ=12) into the TI-30 IIS statistics function, you get:

σ_X =696.49

Compare the above value with the correct answer 699, which we obtained by multiplying probabilities by 10,000. You see that the TI-30X IIS value is about $2 off from the correct answer. Though having $2 off is not far off, in the exam you want to be more precise. It is for this reason that you should not use the TI-30X IIS statistics function.

Which way is better for calculating $E(X)$ and $Var(X)$ of a discrete random variable, using the TI-30X IIS "recalculation" functionality or using the BA II Plus Statistics Worksheet? Both methods are good. I recommend that you learn both methods and choose one you think is better.

Next, let's move on to finding the conditional expectation $E(X|Y)$ and the conditional variance $Var(X|Y)$ of a discrete random variable X.

If an exam problem gives you a list of data pairs of (X,Y) and asks you to find $E(X|Y)$ and $Var(X|Y)$, you can reorganize the original data pairs into new data pairs of $[(X|Y), Prob(X|Y)]$ and then use BA II Plus or TI-30X IIS to calculate the conditional mean and conditional variance.

Problem 3

Two discrete random variables X,Y have the following joint distributions:

		Y = 2	Y = 5	Y = 7	Y = 8
X	0	0.08	0.07	0.09	0.06
	1	0.04	0.1	0.11	0.03
	2	0.01	0.02	0.05	0.12
	3	0.05	0.07	0.04	0.06

Find:

(1) $E(X), Var(X)$

(2) $E(X|Y=5), Var(X|Y=5)$

(3) $Var(XY)$

(4) $Cov(XY)$

(5) $Var\left[\dfrac{X}{Y}\right]$

(6) $Var(X^2Y)$

Solution

(1) $E(X), Var(X)$

Use BA II Plus. To make things easier, sum up the probabilities for each row and for each column:

		Y				
		2	5	7	8	Sum
X	0	0.08	0.07	0.09	0.06	0.30
	1	0.04	0.10	0.11	0.03	0.28
	2	0.01	0.02	0.05	0.12	0.20
	3	0.05	0.07	0.04	0.06	0.22
	Sum	0.18	0.26	0.29	0.27	

For example, in the above table, 0.18=0.08+0.04+0.01+0.05, and 0.3=0.08+0.07+0.09+0.06.

To find $E(X)$ and $Var(X)$, we need to convert the (X,Y) joint distribution table into the X -marginal distribution table:

X	$\Pr(X)$	scaled-up probability $=100 \times \Pr(X)$
0	0.3	30
1	0.28	28
2	0.2	20
3	0.22	22

Next, enter the following data pairs into BA II Plus 1-V Statistics Worksheet: (0,30), (1,28), (2,20), (3,22). Specifically, you enter

X01=0 Y01=30; X02=1 Y02=28;
X03=2 Y03=20; X04=3 Y04=22

Remember in BA II Plus 1-V Statistics Worksheet, you always enter the values of the random variable before entering the scaled-up probabilities. Switching the order will produce a nonsensical result.

After entering the data into BA II Plus, you should get:

$$\overline{X}=1.34, \ \sigma_X=1.1245$$

To find $Var(X) = \sigma_X{}^2$, you do not need to leave the Statistics Worksheet. After BA II Plus displays the value of σ_X, you simply press the x^2 key. This gives you the variance. Please note that squaring σ_X does not change the internal value of σ_X. After squaring σ_X, you can press the up arrow ↑ and the down arrow↓. This removes the squaring effect and gives you the original σ_X.

$$Var(X) = \sigma_X{}^2 = 1.1245^2 = 1.2644$$

So we have $E(X) = 1.34, \quad Var(X) = 1.2644$

(2) $E(X \mid Y = 5), Var(X \mid Y = 5)$

Once again we'll create a table listing all possible values of $(X \mid Y = 5)$ and probabilities of $\Pr(X \mid Y = 5)$:

		$Y = 5$	$\Pr(X \mid Y = 5)$	$100 \times 0.26 \times \Pr(X \mid Y = 5)$
	0	**0.07**	**0.07/0.26**	**7**
X	**1**	**0.10**	**0.10/0.26**	**10**
	2	**0.02**	**0.02/0.26**	**2**
	3	**0.07**	**0.07/0.26**	**7**
	Total	**0.26**		**26**

In the above table, we used the formula

$$\Pr(X \mid Y = 5) = \frac{\Pr(X, Y = 5)}{\Pr(Y = 5)}$$

For example,

$$\Pr(X = 0 \mid Y = 5) = \frac{\Pr(X = 0, Y = 5)}{\Pr(Y = 5)} = \frac{0.07}{0.26}$$

Then we enter the following data into the BA II Plus 1-V Statistics Worksheet:

X01=0 Y01=7; X02=1 Y02=10;
X03=2 Y03=2; X04=3 Y04=7

We have

$$\overline{X} = 1.3462, \ \sigma_X = 1.1416, \ Var(X) = \sigma_X{}^2 = 1.1416^2 = 1.3033$$

So we have $E(X \mid Y=5)=1.3462$, $Var(X \mid Y=5)=1.3033$

You can see that when using the BA II Plus Statistics Worksheet, instead of scaling up $\frac{\Pr(X, Y=5)}{\Pr(Y=5)}$, we can scale up $\Pr(X, Y=5)$ and ignore $\Pr(Y=5)$ because it is a constant. Ignoring $\Pr(Y=5)$, we have:

		$Y=5$	$100 \times \Pr(X \mid Y=5)$
	0	**0.07**	**7**
X	**1**	**0.10**	**10**
	2	**0.02**	**2**
	3	**0.07**	**7**
	Total	**0.26**	**26**

Then we just enter into the BA II Plus 1-V Statistics Worksheet the following:

X01=0 Y01=7; X02=1 Y02=10;
X03=2 Y03=2; X04=3 Y04=7

Ignoring $\Pr(Y=5)$ gives us the same result. This is a great advantage of using BA II Plus. Next time you are calculating the mean and variance of a random variable in BA II Plus, simply ignore any common constant in each probability number. This saves time.

(3) $Var(XY)$
Once again, we need to create a table listing all of the possible values of XY and their corresponding probabilities.

All possible combinations of XY are the boldface numbers listed below:

		Y = 2	Y = 5	Y = 7	Y = 8
	0	**0**	**0**	**0**	**0**
	1	**2**	**5**	**7**	**8**
X	2	**4**	**10**	**14**	**16**
	3	**6**	**15**	**21**	**24**

$\Pr(XY)$ is listed below in boldface:

		Y = 2	Y = 5	Y = 7	Y = 8
	0	**0.08**	**0.07**	**0.09**	**0.06**
X	1	**0.04**	**0.1**	**0.11**	**0.03**
	2	**0.01**	**0.02**	**0.05**	**0.12**
	3	**0.05**	**0.07**	**0.04**	**0.06**

We then scale up $\Pr(XY)$. We multiply $\Pr(XY)$ by 100:

		Y			
		2	5	7	8
X	0	**8**	**7**	**9**	**6**
	1	**4**	**10**	**11**	**3**
	2	**1**	**2**	**5**	**12**
	3	**5**	**7**	**4**	**6**

Next, we enter XY and the scaled-up $\Pr(XY)$ into the BA II Plus 1-V Statistics Worksheet:

X01=0, Y01=8; X02=0, Y02=7; X03=0, Y03=9; X04=0, Y04=6;
X05=2, Y05=4; X06=5, Y06=10; X07=7, Y07=11; X08=8, Y08=3;
X09=4, Y09=1; X10=10, Y10=2; X11=14, Y11=5; X12=16, Y12=12;
X13=6, Y13=5; X14=15, Y14=7; X15=21, Y15=4; X16=24, Y16=6.

Note: you can also consolidate the scaled-up probability for X01=0 into Y01=8+7+9+6=30.

BA II Plus should give you:

$$\overline{X} = 8.0800,\ \sigma_X = 7.5574, Var(X) = \sigma_X{}^2 = 7.5574^2 = 57.1136$$

So we have $E(XY) = 8.08$, $\quad Var(XY) = 57.1136$

Make sure BA II Plus gives you n=100. If n is not 100, at least one of your scaled-up probabilities is wrong.

(4) $Cov(XY)$.

$Cov(XY) = E(XY) - E(X) \times E(Y)$. We already know $E(XY)$ and $E(X)$. We just need to calculate $E(Y)$.

Y	**2**	**5**	**7**	**8**
$\Pr(Y)$	**0.18**	**0.26**	**0.29**	**0.27**
$100\Pr(Y)$	**18**	**26**	**29**	**27**

We enter the following data into the BA II Plus 1-V statistics worksheet:

X01=2, Y01=18; X02=5, Y02=26; X03=7, Y03=29; X04=8, Y04=27;

We should get:

$\overline{X} = 5.8500, \ \sigma_X = 2.1184, \quad Var(X) = \sigma_X{}^2 = 2.1184^2 = 4.4875$

So we have $E(Y) = 5.85, \quad Var(Y) = 4.4875$

Then $\text{Cov}(XY) = E(XY) - E(X) \times E(Y) = 8.08 - 1.34 \times 5.85 = 0.2410$

(5) $Var\left[\frac{X}{Y}\right]$

All of the possible values of $\frac{X}{Y}$ are listed below in boldface:

		Y 2	5	7	8
	0	**0**	**0**	**0**	**0**
X	1	**1/2**	**1/5**	**1/7**	**1/8**
	2	**2/2**	**2/5**	**2/7**	**2/8**
	3	**3/2**	**3/5**	**3/7**	**3/8**

$\Pr\left(\frac{X}{Y}\right)$ are below in boldface:

		Y 2	5	7	8
	0	**0.08**	**0.07**	**0.09**	**0.06**
X	1	**0.04**	**0.1**	**0.11**	**0.03**
	2	**0.01**	**0.02**	**0.05**	**0.12**
	3	**0.05**	**0.07**	**0.04**	**0.06**

The scaled-up probabilities are below in boldface:

		2	Y 5	7	8
	0	**8**	**7**	**9**	**6**
X	1	**4**	**10**	**11**	**3**
	2	**1**	**2**	**5**	**12**
	3	**5**	**7**	**4**	**6**

Enter the following into the BA II Plus 1-V Statistics Worksheet:

X01=0, Y01=8; X02=0, Y02=7; X03=0, Y03=9; X04=0, Y04=6;
X05=1/2, Y05=4;X06=1/5, Y06=10; X07=1/7, Y07=11; X08=1/8, Y08=3;
X09=2/2, Y09=1;X10=2/5, Y10=2;X11=2/7, Y11=5;X12=2/8, Y12=12;
X13=3/2, Y13=5;X14=3/5, Y14=7;X15=3/7, Y15=4;X16=3/8, Y16=6.

You should get:

$$\bar{X} = 0.2784, \quad \sigma_X = 0.3427$$

This means that

$$E\left[\frac{X}{Y}\right] = 0.2784, \quad Var\left[\frac{X}{Y}\right] = 0.1175$$

(6) $Var(X^2Y)$ --- This problem is for you to solve.

Answer: $E(X^2Y) = 18.2, \quad Var(X^2Y) = 477.2$.

Please note in solving (3),(4),(5),(6), you are using the same scaled-up probabilities. You can reuse the scaled-up probabilities you entered for the previous problem. You just enter the new series of X values in the BA II Plus 1-V Statistics Worksheet.

Though tedious, this problem is a good opportunity for you to practice how to find the mean and variance of a discrete random variable using the BA II Plus Statistics Worksheet.

Homework for you: Sample P #114. Hint: Enter X01=0, Y01=50; X02=1, Y02=125.

Chapter 12 Bernoulli distribution

If X is a Bernoulli random variable, then

$$X = \begin{cases} 1 & \text{with probability of } p \ (0 \le p \le 1) \\ 0 & \text{with probability of } q = 1 - p \end{cases}$$

Examples:

- The single flip of a coin resulting in heads or tails
- The single throw of a die resulting in a 6 or non-6 (1 through 5)
- A single contest resulting in winning or losing
- The launch of a product resulting in either success or failure

Key formulas:

$$E(X) = p, \ Var(X) = pq, \ M_X(t) = q + pe^t$$

Sample Problems and Solutions

Problem 1

During one policy year, a certain life insurance policy either results in no claims with a probability of 99% when the policyholder is still alive or results in a single claim if the policyholder dies.

Let X represent the number of claims on the policy during one policy year. Find $E(X)$, σ_X, and $M_X(t)$.

Solution

X is a Bernoulli random variable with $p = 1\%$

$$E(X) = p = 1\%, \ \sigma_X = \sqrt{Var(X)} = \sqrt{pq} = \sqrt{1\%(99\%)} = 9.95\%$$

$$M_X(t) = q + pe^t = 0.99 + 0.01e^t$$

Homework for you – rework all the problems listed in this chapter

Chapter 13 Binomial distribution

Binomial distribution is one of the most frequently used and most intuitive distributions in Exam P. The easiest way to understand Binomial distribution is to think about tossing coins. You toss the same coin *n* times and you want to know how many times you can get heads. For each toss, there are only two outcomes -- either you get heads or tails. The probability of getting heads, represented by *p*, doesn't change from one toss to another. In addition, whether you get a head in one toss doesn't affect whether you get heads in the next toss (so the result of each toss is independent). If you let X = # of heads you get in *n* tosses of the coin, then X has a binomial distribution.

Key formulas to remember:

For a binomial random variable X

$$p_X(x) = C_n^x p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}, \ 0 \le p \le 1, \ q = 1-p, \ x = 0,1,2,...,n$$

$$E(X) = np, \ Var(X) = npq, \ M_X(t) = (q + pe^t)^n$$

Let's get an intuitive feel for the formula $p_X(x) = C_n^x p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$. If there is only one way of having x successes and $(n-x)$ of failures, then the probability of having exactly x successes is $p^x q^{n-x}$ due to the independence among n trials. However, out of n trials, there are $\frac{n!}{x!(n-x)!}$ ways of having x successes and $n-x$ of failures. So the total probability of having exactly x successes is $\frac{n!}{x!(n-x)!} p^x q^{n-x}$.

Shortcuts

<u>Shortcut #1 --- How to quickly find the probability mass function of a binomial distribution for *n*=2 and *n*=3 (which is often tested).</u>

You can easily memorize the probability mass function for n=2 and n=3, eliminating the need to derive the probability from the generic formula

$$p_X(x) = C_n^x p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

To memorize the probability mass function for *n*=2, notice that

$$1 = (p+q)^2 = p^2 + 2pq + q^2 .$$

In the above formula, each item on the right-hand size represents the probability mass function for n=2. In other words,

$$p_X(2) = p^2, \; p_X(1) = 2pq, \; p_X(0) = q^2$$

Similarly, to quickly find the probability mass function for *n*=3, notice

$$1 = (p+q)^3 = p^3 + 3p^2q + 3pq^2 + q^3$$

Then $p_X(3) = p^3$, $p_X(2) = 3p^2q$, $p_X(1) = 3pq^2$, $p_X(0) = q^3$

Shortcut #2 – How to memorize the mean, variance, and moment generating function of a binomial variable.

A binomial random variable X is simply the sum of n independent identically distributed Bernoulli variable Y . Y has the following probability mass function:

$$Y = \begin{cases} 1 & \text{with probability of } p \quad (0 \le p \le 1) \\ 0 & \text{with probability of } q = 1 - p \end{cases}$$

The mean and variance of Y are:

$$E(Y) = p, \qquad Var(Y) = pq$$

Since $X = Y_1 + Y_2 + \ldots + Y_n$ and $Y_1, Y_2, \ldots, Y_n$ are independent identically distributed Bernoulli variables,

$$E(X) = E(Y_1 + Y_2 + \ldots + Y_n) = nE(Y) = np$$
$$Var(X) = Var(Y_1 + Y_2 + \ldots + Y_n) = nVar(Y) = npq$$
$$M_X(t) = M_{Y_1+Y_2+\ldots+Y_n}(t) = \left[M_Y(t)\right]^n = \left(q + pe^t\right)^n$$

The last equation is based on the rule that the MGF (moment generating function) of the sum of independent random variables is simply the product of the MGF's of each individual random variable.

The mean $E(X) = np$ should make sense to you. In one trial, the number of successes you can expect should be p, because p is the probability of having a success in one trial. Then if you have n independent trials, you should expect to have a total of np successes.

Application of binomial distribution in insurance

Binomial distribution is frequently used to model the frequency (the number of losses), not the severity (the dollar amount of individual losses), of a random loss variable. For example, the total number of burglaries or fires or some other random event that happens to a group of geographically dispersed homes is approximately a binomial random variable.

You can easily see that the assumptions underlying a binomial random variable approximately hold in this case. Because fire is a rare event, you can assume that for a given period of time (6 months or 1 year for example), a house has no fire or only 1 fire. You can also assume that the probability of having a fire is constant from one city to another (to keep your model simple). Since the houses you are studying are located in different areas, it is reasonable to assume that whether one house has a fire does not affect the likelihood of another house catching fire. Then you can use the probability mass function of binomial distribution to calculate the total number of fires that can happen to a block of houses for a given period of time.

Determine whether you have a binomial random variable.

Interestingly, if the probability distribution is not binomial (Poisson, exponential, or some other non-binomial related distribution), the problem will most likely give you the actual distribution ("The number of claims filed has a Poisson distribution…"). However, if the distribution is binomial, negative binomial, or geometric, the exam question most likely won't tell you so and you have to figure out the distribution yourself. This is because binomial, negative binomial, or geometric distributions can often be identified unambiguously from the context of a tested problem. In contrast, Poisson, exponential, and many other non-binomial related distributions are not self-evident; SOA may need to give you the actual distribution to avoid possible ambiguity around a tested problem.

To verify whether a distribution is indeed binomial, you can check whether the random variable meets the following standards:

1. There are n independent trials. The outcome of one trial does not affect the outcome of any other trial.

2. Each trial has only two outcomes – success or failure.

3. The probability of success remains constant throughout n trials.

4. You are determining the probability of having x number of successes, where x is a non-negative integer no greater than n.

Once you have identified that you have a binomial random variable, just plug data into the memorized formula for the probability mass function, the mean, and the variance.

Common Pitfalls to Watch for

Some candidates erroneously drop the combination factor $C_n^x = \dfrac{n!}{x!(n-x)!}$ when calculating the probability mass function of a binomial random variable. For example, to find the probability of having one success in two independent trials, some candidates write

$$p_X(1) = pq$$

Their logic is this: the probability of having one success in one trial is p and the probability of having one failure in another trial is q. Hence the probability of having one success and one failure is $p_X(1) = pq$ according to the multiplication rule.

The above logic is wrong because there are two ways to have exactly one success out of two independent trials. One way is to succeed in the first trial and fail the second trial. The other way is to fail the first trial and succeed in the second trial. So the correct formula for $n = 2$ is $p_X(1) = 2pq$.

Sample Problems and Solutions

Problem 1

On a given day, each computer in a lab has at most one crash. There is a 5% chance that a computer has a crash during the day, independent of the performance of any other computers in the lab. There are 25 computers in the lab.

Find the probability that on a given day, there are
(1) exactly 3 crashes
(2) at most 3 crashes
(3) at least 3 crashes
(4) more than 3 and less than 6 crashes

Solution

X = # of computer crashes in a given day. X has a binomial distribution with $n = 25$ and p =5%.

$$p_X(x) = C_n^x p^x q^{n-x} = C_{25}^x (5\%)^x (95\%)^{25-x}$$

Probability of having exactly 3 crashes in a day is:
$p_X(3) = C_{25}^{3}(5\%)^{3}(95\%)^{25-3}$ =9.3%

Probability of having at most 3 crashes in a day is:
$$P(X \leq 3) = p_X(0) + p_X(1) + p_X(2) + p_X(3) = \sum_{x=0}^{3} C_{25}^{x}(5\%)^{x}(95\%)^{25-x}$$
$$= C_{25}^{0}(5\%)^{0}(95\%)^{25-0} + C_{25}^{1}(5\%)^{1}(95\%)^{25-1} + C_{25}^{2}(5\%)^{2}(95\%)^{25-2} + C_{25}^{3}(5\%)^{3}(95\%)^{25-3}$$
=96.6%

Probability of having at least 3 crashes in a day is:
$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \sum_{x=0}^{2} C_{25}^{x}(5\%)^{x}(95\%)^{25-x}$$
$$= 1 - \left[C_{25}^{0}(5\%)^{0}(95\%)^{25-0} + C_{25}^{1}(5\%)^{1}(95\%)^{25-1} + C_{25}^{2}(5\%)^{2}(95\%)^{25-2}\right]$$
=12.7%

Alternatively, the probability of having at least 3 crashes in a day is:

$$P(X \geq 3) = 1 - P(X \leq 3) + P(X = 3) = 1 - P(\text{at most 3 crashes}) + P(\text{3 crashes})$$
= 1-96.6%+9.3%=12.7%

Probability of having more than 3 but less than 6 crashes in a day is:
$p_X(4) + p_X(5) = C_{25}^{4}(5\%)^{4}(95\%)^{25-4} + C_{25}^{5}(5\%)^{5}(95\%)^{25-5}$ =3.3%

Problem 2

A factory has 25 machines working separately, of which 15 are of Type A and 10 of Type B. On a given day, each machine Type A has a 5% chance of malfunctioning, independent of the performance of any other machines; each machine Type B has an 8% chance of malfunctioning, independent of the performance of any other machines.

Let Y represent, on a given day, the total number of machines malfunctioning.

Find $P(Y = 2)$, the probability of having exactly 2 machines malfunctioning on a given day.

Solution

On a given day, the number of Type A machines malfunctioning X_A and the number of Type B machines malfunctioning X_B are two independent binomial random variables. X_A is binomial with n=15 and p=5%; X_B is binomial with n=10 and p=8%.

$Y = X_A + X_B$. There are only 3 combinations of X_A and X_B to make $Y = 2$: two Type A's malfunction and no Type B's malfunction; one Type A and one Type B malfunction; and no Type A's malfunction and two Type B's malfunction. In other words,

$$Y = X_A + X_B = 2 \Leftrightarrow (X_A, X_B) = (2,0), (1,1), (0,2)$$

We just need to find the probability of each of the above three combinations and calculate the sum.

$$P(X_A = x_A, X_B = x_B) = P(X_A = x_A)P(X_B = x_B)$$

Because X_A and X_B are independent

$$P(X_A = 2, X_B = 0) = P(X_A = 2)P(X_B = 0)$$
$$P(X_A = 2) = C_{15}^{2}(5\%)^2(95\%)^{15-2} = 13.48\%$$
$$P(X_B = 0) = C_{10}^{0}(8\%)^0(92\%)^{10-0} = 43.44\%$$
$$\Rightarrow \quad P(X_A = 2, X_B = 0) = 13.48\% * 43.44\% = 5.85\%$$

$$P(X_A = 1, X_B = 1) = P(X_A = 1)P(X_B = 1)$$
$$= C_{15}^{1}(5\%)^1(95\%)^{15-1}C_{10}^{1}(8\%)^1(92\%)^{10-1} = 13.82\%$$

$$P(X_A = 0, X_B = 2) = P(X_A = 0)P(X_B = 2)$$
$$= C_{15}^{0}(5\%)^0(95\%)^{15-0}C_{10}^{2}(8\%)^2(92\%)^{10-2} = 6.85\%$$

$$\Rightarrow \quad P(Y = 2) = 5.85\% + 13.82\% + 6.85\% = 26.52\%$$

Problem 3 (CAS Exam 3 Spring 2005 #2, wording simplified)

BIB is new insurer writing homeowner policies. You are given:

- BIB's initial wealth in a saving account = $15
- Number of insured homes = 3
- Annual premium of $10 per homeowner policy collected at the beginning of each year
- Annual claim per homeowner policy, if any, is $40 and is paid immediately
- Beside claims, BIB doesn't have any other expenses
- BIB's saving account earns zero interest

Each homeowner files at most one claim per year. The probability that each homeowner files a claim in Year 1 is 20%. The probability that each homeowner files a claim in Year 2 is 10%. Claims are independent.

Calculate the probability that BIB will NOT go bankrupt in the first 2 years.

Solution

BIB will not go bankrupt at in the first 2 years (i.e. BIB still stays in business at the end of Year 2) if there is at most 1 claim in the first 2 years.

If there is at most 1 claim in the 1st two years, then

BIB's total wealth at the end of Year 2 before paying any claims
= initial wealth
+ premiums collected during Year 1
+ premiums collected during Year 2
= \$15 + \$10 (3) + \$10 (3) = \$75

BIB's total expense incurred during Year 1 and Year 2
= cost per claim $\times$ total # of claims during the first two years
= \$40 (1) = \$40

BIB's total wealth at the end of Year 2 after paying the claims
= \$75 - \$40 = \$35

If BIB incurs two claims in the first two years, then the claim cost during the first two years is \$40(2)=\$80. BIB's total wealth at the end of Year 2 is \$75 - \$80 = - \$5. BIB will go bankrupt at the end of Year 2.

So BIB needs to have zero or one claim in the first two years to stay in business. There are three ways to have zero or one claim during the first two years:

- Have zero claim in Year 1 and Year 2 (Option 1)
- Having zero claim in Year 1 and 1 claim in Year 2 (Option 2)
- Having one claim in Year 1 and zero claim in Year 2 (Option 3)

The number of claims in Year 1 is a binominal distribution with parameter $n = 3$ and $p = 0.2$. The number of claims in Year 2 is a binominal distribution with parameter $n = 3$ and $p = 0.1$.

Next, we set up a table keeping track of claims:

			A	B	C=A×B
Option	Year 1	Year 2	Yr 1 Probability	Yr 2 Probability	Total Probability
#1	0	0	$C_3^0\left(0.2^0\right)\left(0.8^3\right)$ $=0.8^3$	$C_3^0\left(0.1^0\right)\left(0.9^3\right)$ $=0.9^3$	$0.8^3\left(0.9^3\right)$ $=37.32\%$
#2	0	1	$C_3^0\left(0.2^0\right)\left(0.8^3\right)$ $=0.8^3$	$C_3^1\left(0.1^1\right)\left(0.9^2\right)$ $=3(0.1)0.9^2$	$0.8^3(3)(0.1)0.9^2$ $=28\%$
#3	1	0	$C_3^1\left(0.2^1\right)\left(0.8^2\right)$ $=3(0.2)\left(0.8^2\right)$	$C_3^0\left(0.1^0\right)\left(0.9^3\right)$ $=0.9^3$	$3(0.2)\left(0.8^2\right)\left(0.9^3\right)$ $=12.44\%$
Total					77.76%

The probability is 77.76% that BIB will not go bankrupt during the first 2 years.

Problem 4 (CAS Exam 3 Spring 2005 #15)

A service guarantee covers 20 TV sets. Each year, each set has 5% chance of failing. These probabilities are independent.

If a set fails, it is immediately replaced with a new set at the end of the year of failure. This new set is included in the service guarantee.
Calculate the probability of no more than one failure in the first two years.

Solution

There are three ways to have zero or one failures during the first two years:

- Have zero failure in Year 1 and Year 2 (Option 1)
- Having zero failure in Year 1 and 1 failure in Year 2 (Option 2)
- Having one failure in Year 1 and zero failure in Year 2 (Option 3)

The # of failures each year is a binomial distribution with parameter $n=20$ and $p=5\%$.

Next, we set up a table keeping track of claims:

			A	B	C=A×B
	Yr1	Yr 2	Yr 1 Probability	Yr 2 Probability	Total Probability
#1	0	0	$C_{20}^{0}\left(0.05^{0}\right)\left(0.95^{20}\right)$ $=0.95^{20}$	$C_{20}^{0}\left(0.05^{0}\right)\left(0.95^{20}\right)$ $=0.95^{20}$	$\left(0.95^{20}\right)^{2}$ =12.85%
#2	0	1	$C_{20}^{0}\left(0.05^{0}\right)\left(0.95^{20}\right)$ $=0.95^{20}$	$C_{20}^{1}\left(0.05^{1}\right)\left(0.95^{19}\right)$ $=20\left(0.05\right)0.95^{19}$	$20\left(0.05\right)0.95^{39}$ =13.53%
#3	1	0	$C_{20}^{1}\left(0.05^{1}\right)\left(0.95^{19}\right)$ $=20\left(0.05\right)0.95^{19}$	$C_{20}^{0}\left(0.05^{0}\right)\left(0.95^{20}\right)$ $=0.95^{20}$	$20\left(0.05\right)0.95^{39}$ =13.53%
Total					39.91%

The probability of no more than one failure in the first two years is 39.91%.

Homework for you:
#40 May 2000; #23, #37 May 2001; #27 Nov 2001.

Chapter 14 Geometric distribution

You perform Bernoulli trials (flipping a coin, throwing a die) repeatedly till you get your first success (for example -- getting a first head or first 6), then you stop and count the total number of trials you had (including the final trial that brings you the only success). The total number of trials you counted is a geometric random variable.

If we let X = # of independent **trials** just enough for one success, and p is the constant probability of success in any single trial, then X is geometric random variable with parameter p.

Please note that one nasty thing about a geometric (and a negative binomial) random variable is that there are always two ways to define the random variable – to define it as the number of trials or the number of failures. Though the variance formula is identical either way, the mean formulas are different (the two means differ by one).

Because exam questions can use either way to define a geometric variable, you need to determine, according to the context of the problem, which mean and probability mass formulas to use.

Here is a second way to define a geometric random variable:
If we let Y = # of independent **failures** just enough for one success, p is the constant probability of success in any single trial, then Y is geometric random variable with parameter p.

Let's look at the probability mass function of X and Y :

S=success, F=failure, p=the probability of success in a single trial. Then

X = # of trials before 1st success	1	2	3	4	…	n
Y = # of failures before 1st success	0	1	2	3		k $(k=n-1)$
Outcome in terms of trials	S	FS	FFS	FFFS	…	$\underbrace{FF...F}_{n-1}S$
Outcome in terms of failures	S	FS	FFS	FFFS	…	$\underbrace{FF...F}_{k}S$
Probability $P(X=n)$	p	$(1-p)p$	$(1-p)^2 p$	$(1-p)^3 p$	…	$(1-p)^{n-1} p$
Probability $P(Y=k)$	p	$(1-p)p$	$(1-p)^2 p$	$(1-p)^3 p$	…	$(1-p)^k p$

Key formulas for geometric random variable

	Formula	Explanation
PMF (probability mass function)	$p_X(n)=(1-p)^{n-1}p$ $p_X(k)=(1-p)^k p$	To have 1st success at *n*-th trial, you must fail the first (*n*-1) trials and succeed at n-th trial. Notice *X*=*Y*+1.
$P(X\geq n)$ $P(Y\geq k)$	$P(X\geq n)=(1-p)^{n-1}$ $P(Y\geq k)=P(X\geq k+1)=(1-p)^k$	To need at least n trials to get one success, you must have zero success in the first $n-1$ trials.
Cumulative probability mass function	$F_X(n)=P(X\leq n)=1-P(X\geq n+1)$ $=1-(1-p)^n$ $F_Y(k)=P(Y\leq k)=1-P(Y\geq k+1)$ $=1-(1-p)^{k+1}$	The continuous counterpart of geometric is exponential. Exponential CDF is $F_X(x)=1-e^{-\lambda x}$. Notice that CDF for geometric and exponential is one minus something.
Mean	$E(X)=\dfrac{1}{p}$ $E(Y)=E(X)-1=\dfrac{1}{p}-1$	You can derive it using MGF (next page). Intuitive formula. For example, if $p=0.2=1/5$, then on average every 5 trials bring in one success. So $E(X)=1/p$.
Variance	$Var(X)=Var(Y)=\dfrac{1}{p^2}-\dfrac{1}{p}$	You can derive it using MGF (next page).
MGF	$M_X(t)=\dfrac{pe^t}{1-(1-p)e^t}$ $M_Y(t)=M_{X-1}(t)=M_X(t)e^{-t}=\dfrac{p}{1-(1-p)e^t}$	Not an intuitive formula. Just have to memorize the first one. To get the 2nd formula, use $M_{X+c}(t)=M_X(t)e^{ct}$
Conditional probability (**Lack of memory property**)	$P(X\geq a+b \mid X\geq b+1)$ $=\dfrac{P(X\geq a+b)}{P(X\geq b+1)}=\dfrac{(1-p)^{a+b-1}}{(1-p)^b}$ $=(1-p)^{a-1}=P(X\geq a)$ $P(Y\geq a+b \mid Y\geq b)$ $=\dfrac{P(Y\geq a+b)}{P(Y\geq b)}=\dfrac{(1-p)^{a+b}}{(1-p)^b}$ $=(1-p)^a=P(Y\geq a)$	X and Y don't remember their past. After making initial b trials with no success, your chance of succeeding after at least a more trials is the same as if you would reset your counter to zero after the initial b trials and start counting trials afresh. Why so? If an event is truly random (such as coin tossing, your past success or failure should have no bearing on your future success or failure.

How to derive the mean and variance formula using MGF:

$$M_X(t)=\frac{pe^t}{1-(1-p)e^t}=\frac{p}{e^{-t}-(1-p)}$$

$$\frac{d}{dt}M_X(t)=p\frac{d}{dt}\left[e^{-t}-(1-p)\right]^{-1}=pe^{-t}\left[e^{-t}-(1-p)\right]^{-2}$$

$$\Rightarrow \qquad E(X)=\left[\frac{d}{dt}M_X(t)\right]_{t=0}=\left\{pe^{-t}\left[e^{-t}-(1-p)\right]^{-2}\right\}_{t=0}=p\left[1-(1-p)\right]^{-2}=\frac{1}{p}$$

$$\frac{d^2}{dt^2}M_X(t)=\frac{d}{dt}\left\{pe^{-t}\left[e^{-t}-(1-p)\right]^{-2}\right\}$$
$$=-pe^{-t}\left[e^{-t}-(1-p)\right]^{-2}+pe^{-t}(-2)\left(-e^{-t}\right)\left[e^{-t}-(1-p)\right]^{-3}$$

$$E\left(X^2\right)=\left[\frac{d^2}{dt^2}M_X(t)\right]_{t=0}=\left\{-pe^{-t}\left[e^{-t}-(1-p)\right]^{-2}+pe^{-t}(-2)\left(-e^{-t}\right)\left[e^{-t}-(1-p)\right]^{-3}\right\}_{t=0}$$
$$=-p\left[1-(1-p)\right]^{-2}+2p\left[1-(1-p)\right]^{-3}=-p\left(p^{-2}\right)+2p\left(p^{-3}\right)=\frac{2}{p^2}-\frac{1}{p}$$

$$\Rightarrow \qquad Var(X)=E\left(X^2\right)-E^2(X)=\left(\frac{2}{p^2}-\frac{1}{p}\right)-\frac{1}{p^2}=\frac{1}{p^2}-\frac{1}{p}$$

Sample Problems and Solutions

Problem 1

A candidate is taking a multiple-choice exam. For each tested problem in the exam, there are 5 possible choices --- A, B, C, D, and E. Because the candidate has zero knowledge of the subject, he relies on pure guesswork to answer each question, independent of how he answers any previous questions.

Find:

(1) The probability that the candidate answers three questions wrong in a row before he finally answers the fourth question correctly.

(2) Let random variable X = # of problems the candidate answers wrong in a row before he finally guesses a correct answer. Find $E(X)$ and $Var(X)$.

(3) When the candidate's exam is being graded, the grader finds that the candidate has answered the first five exam questions all wrong. Can the grader conclude that the candidate is more or less likely to guess a problem right in other problems?

(4) When the candidate's exam is being graded, the grader finds that the candidate has answered the first two exam questions all correctly. Can the grader conclude that the candidate is more or less likely to guess a problem right in other problems?

Solution

(1) The probability of having three wrongs but the fourth right.

If you use the number of failures X as the geometric random variable, then you have

$p_X(x) = p(1-p)^x$, x =0,1,2,3,… and p =0.2

$p_X(3) = p(1-p)^3 = 0.2(1-0.2)^3$ =10.24%

If you use the number of trials Y as the geometric random variable, then you have:

$p_Y(y) = p(1-p)^{y-1}$, y =1,2,3,… and p =0.2

$p_Y(4) = p(1-p)^{4-1} = 0.2(1-0.2)^3$ =10.24%

(2) Find $E(X)$ and $Var(X)$

If you use the number of failures X as the geometric random variable, then you have

$$E(X) = \frac{1}{p} - 1 = \frac{1}{0.2} - 1 = 4,\ Var(X) = \frac{1}{p^2} - \frac{1}{p} = \frac{1}{0.2^2} - \frac{1}{0.2} = 20$$

If you use the number of trials Y as the geometric random variable, then you have

$$E(Y) = \frac{1}{p} = \frac{1}{0.2} = 5,\ Var(Y) = \frac{1}{p^2} - \frac{1}{p} = \frac{1}{0.2^2} - \frac{1}{0.2} = 20$$

(3) Geometric distribution lacks memory. The past failures have no use in predicting future successes or failures.

If the candidate really has zero knowledge of the subject and all exam problems are unrelated, then the candidate answering the first 5 questions all wrong is purely by chance. This information is useless for predicting the candidate's performance on other exam questions.

(4) Once again, geometric distribution lacks memory. The past successes have no bearing on future successes or failures.

Problem 2

You throw a die repeatedly until you get a 6. What's the probability that you need to throw more than 20 times to get 6?

Solution

If you use the number of trials X as the geometric random variable, then you have

$$P(X \geq n) = (1-p)^{n-1}, \ p = 1/6$$
$$P(X > 20) = P(X \geq 21) = (1-p)^{21-1} = (1-1/6)^{21-1} = 2.6\%$$

If you use the number of failures Y as the geometric random variable, then you have:
$P(Y \geq k) = (1-p)^k$

Throwing a die at least 21 times to get a 6 is the same as having at least 20 failures before a success. $P(Y \geq 20) = (1-p)^{20} = (1-1/6)^{20}$ =2.6%

Problem 3

The annual number of losses incurred by a policyholder of an auto insurance policy, N , is geometrically distributed with parameter p =0.8. If losses do occur, the amount of losses is either $1,000 with probability of 0.4 or $5,000 with probability of 0.6. Assume the number of losses and amounts of losses are independent.

Let S =annual aggregate loss incurred by the policyholder.

Find $E(S)$ and $Var(S)$.

Solution

Let X =amount of a single loss (set K= $1,000 to make our calculation easier):

$$X = \begin{cases} 1K & \text{with probability } 0.4 \\ 5K & \text{with probability } 0.6 \end{cases}$$

We wrote $1K$ and $5K$ so we won't forget that the unit is $1,000.

$S = NX$

Because N and X are independent, we have

$E(S) = E(N)E(X)$ (intuitive)

$Var(S) = Var(NX) = Var(N)E^2(X) + E(N)Var(X)$ (not intuitive)

You should memorize the above formulas.

Note in the second formula above, the first item on the right hand side is $Var(N)E^2(X)$, not $Var(N)E(X)$. To see why, notice that $Var(NX)$ is $\$^2$ (variance of a dollar amount is dollar squared); $E(N)Var(X)$ is dollar squared. $Var(N)E^2(X)$ is also dollar squared. If you use $Var(N)E(X)$ as opposed to the correct term $Var(N)E^2(X)$, then $Var(N)E(X)$ is a dollar amount, which will not make the equation hold.

To calculate $E(X)$ and $Var(X)$, we simply plug in the following data (remember to scale up the probability) into BA II Plus (by now this should be your second nature):

$$X = \begin{cases} 1K & \text{with probability } 0.4 \\ 5K & \text{with probability } 0.6 \end{cases}$$

Set X01=1, Y01=4; X02=5, Y02=6. Using BA II Plus 1-V Statistics Worksheet, you should get:

$$E(X) = 3.4K \text{ and } Var(X) = 3.84K^2$$

Alternatively,

$$E(X) = 1(0.4) + 5(0.6) = 3.4\text{K}$$

$$E(X^2) = 1^2(0.4) + 5^2(0.6) = 15.4K^2$$

$$Var(X) = E(X^2) - E^2(X) = 15.4 - 3.4^2 = 3.84K^2$$

Next, we need to find the mean and variance of N. N has a geometric distribution with parameter p =0.8. Should we treat N as the number of trials or the number of failures? Because N is the number of losses occurred in a year and a policyholder may have zero losses in a year, N starts from zero, not from one. We should treat N as the number of failures and use the mean formula that produces the lower mean:

$$E(N) = \frac{1}{p} - 1 = \frac{1}{0.8} - 1 = 0.25, \qquad Var(N) = \frac{1}{p^2} - \frac{1}{p} = \frac{1}{0.8^2} - \frac{1}{0.8} = 0.3125$$

$$E(S) = E(N)E(X) = 3.4K(0.25) = 0.85K = 850$$

$$Var(S) = Var(N)E^2(X) + E(N)Var(X) = 0.3125(3.4K)^2 + 0.25(3.84K^2)$$
$$= 4.5725K^2 = 4.5725(1,000^2) = 4,572,500(\$^2)$$

Problem 4

The annual number of losses incurred by a policyholder of an auto insurance policy, N, is geometrically distributed with parameter p =0.4.

Find $E(N-2|N>2)$, $Var(N-2|N>2)$, $E(N|N>2)$, $Var(N|N>2)$

Solution

Geometric random variable N doesn't have any memory of its past. After seeing k losses (k is a non-negative integer), if we set the counter to zero and start counting the number of losses incurred from this point on (that is $N-k$), the count $N-k$ will have the identical geometric distribution with the same parameter p.

In this problem, after seeing two losses, if we reset the counter to zero, the number of losses we will see in the future is $N-2$, where N is the original count of losses before the counter is reset to zero. $N-2$ has a geometric distribution with the identical parameter p =0.4. In other words, the conditional random variable $N-2|N>2$ is also geometrically distributed with p =0.4.

$$E(N-2|N>2) = E(N) = \frac{1}{p} - 1 = \frac{1}{0.4} - 1 = 1.5$$
$$Var(N-2|N>2) = Var(N) = \frac{1}{p^2} - \frac{1}{p} = \frac{1}{0.4^2} - \frac{1}{0.4} = 3.75$$

$$E(N|N>2) = E(N-2|N>2) + 2 = 1.5+2=3.5$$
$$Var(N|N>2) = Var(N-2|N>2) = 3.75$$

Homework for you: rework all the problems in this chapter.

Chapter 15 Negative binomial

Negative binomial distribution is one of the more difficult concepts in Exam P. And it is very easy for candidates to miscalculate probabilities related to a negative binomial distribution. To understand the negative binomial distribution, remember the following critical points.

Critical Point #1- Negative binomial distribution is like a "negative" version of our familiar binomial distribution. In a binomial distribution, the number of trials n is fixed and we want to find the probability of having k number of successes in these n trials (so K is the random variable). In contrast, in a negative binomial distribution, the number of successes, k, is fixed and we want to find out the probability that these k successes are produced in n trials (so N is the random variable).

Critical point #2 – the probability mass function of a negative binomial distribution with parameter (n,k,p) is a fraction of the probability mass function of the binomial distribution with identical parameters (n,k,p), where $n=$ # of trials, k = # of successes, and p =the probability of success. The fraction is equal to k/n. Expressed in an equation,

$$f^{NB}(n,k,p)=\frac{k}{n}f^{B}(n,k,p),$$ where NB=negative binomial, B=binomial

Stated differently,

$$P(k\text{ successes are produced by } n \text{ trials})=\frac{k}{n}\times P(k\text{ successes in } n\text{ trials})$$

We can easily prove the above equation.

$f^{NB}(n,k,p)$
= P(k successes are produced by n trials)
= P (first n-1 trials produce k-1 successes and final trial is a success)
= P(first n-1 trials produce k-1 successes)× P (final trial is a success)

The above equation stands because each trial is independent.

(first n-1 trials produce k-1 successes) is a binomial distribution; its probability is
$C_{n-1}^{k-1}p^{k-1}(1-p)^{(n-1)-(k-1)}=C_{n-1}^{k-1}p^{k-1}(1-p)^{n-k}$

P (final trial is a success) = p

$$\Rightarrow \quad f^{NB}(n,k,p)=pC_{n-1}^{k-1}p^{k-1}(1-p)^{n-k}=C_{n-1}^{k-1}p^{k}(1-p)^{n-k}$$

However, $C_{n-1}^{k-1}=\dfrac{(n-1)!}{(k-1)!(n-k)!}=\dfrac{k}{n}\times\dfrac{n!}{k!(n-k)!}=\dfrac{k}{n}C_n^k$

$$\Rightarrow \qquad f^{NB}(n,k,p)=C_{n-1}^{k-1}p^k(1-p)^{n-k}=\frac{k}{n}C_n^k p^k(1-p)^{n-k}=\frac{k}{n}f^B(n,k,p)$$

We want to express the negative binomial probability mass function (PMF) as a fraction of the binomial PMF. Binomial is one of the easiest probability functions for candidates to understand and calculate. Since you can find binomial PMF with parameters (n,k,p), you can easily calculate the negative binomial PMF by simply taking a fraction (*k/n*) of the binomial PMF with the identical parameters. This way, you don't have to memorize the negative binomial PMF. You just need to memorize the fraction (*k/n*).

Without using the rigorous proof above, we can still intuitively see that, given the identical parameters (n,k,p), the negative binomial PMF is smaller than the binomial PMF. This is because the negative binomial distribution has more constraints than does the binomial distribution. Both distributions have *k* successes in *n* trials; the PMF for each distribution is a multiple of $p^k(1-p)^{n-k}$. However, the negative binomial distribution requires that

(1) the first $n-1$ trials have k -1 successes,
(2) the final trial, the *n*-th trial, be a success,

In contrast, the binomial distribution does not have these two constraints. These additional constraints reduce the number of ways of having *k* successes in *n* trials in the negative binomial distribution. Mathematically, it happens to be that the number of ways of having *k* successes in *n* trials in a negative binomial distribution is only a fraction (*k/n*) of the number of ways of having *k* successes in *n* trials.

Critical point #3 – A negative binomial distribution with parameters (n,k,p) is the sum of k independent identically distributed geometric distribution with parameter p. If you understand this, memorizing the formulas for the mean and variance of a negative binomial distribution becomes easy.

Let's look into this further. Let

N_1 = # of trials required for 1st success; N_1 is a geometric variable with parameter *p*.

N_2 = # of trials required for 2nd success; N_2 is a geometric variable with parameter p.

……

N_k = # of trials required for *k*-th success, then N_k is a geometric variable with parameter *p*.

And

N = # of trials required for a total of *k* successes. N is a negative binomial distribution with parameter *p*.

Then $N = N_1 + N_2 + ... + N_k$

Because each trial is independent, then N_1, N_2, …, N_k are independent.

$$\Rightarrow \qquad E(N) = E(N_1 + N_2 + ... + N_k) = k \times \frac{1}{p} = \frac{k}{p}$$

$$Var(N) = Var(N_1 + N_2 + ... + N_k) = k\left(\frac{1}{p^2} - \frac{1}{p}\right)$$

To obtain the moment generating function of a negative binomial distribution, we simply raise the moment generating function of a geometric distribution to the power of k:

$$M_N(t) = M_{N_1+N_2+...+N_k}(t) = \left[M_{N_1}(t)\right]^k = \left[M_{N_2}(t)\right]^k = ... = \left[M_{N_k}(t)\right]^k = \left[\frac{pe^t}{1-(1-p)e^t}\right]^k$$

How is the negative binomial distribution used in insurance? It turns out that if

- the number of claims by the insured has a Poisson distribution with mean λ

- λ changes from one risk group to another (i.e. low risk insured's have a lower average claim than a high risk group) and has a gamma distribution,

then the number of claims is a negative binomial distribution. This is called the Gamma-Poisson Model, which is routinely tested in SOA Exam C. The Gamma-Poisson Model is beyond the scope of Exam P. So you do not need to worry about this. As you progress in your actuarial career and exams, you will pick up this concept. For now just keep in mind that the number of claims by the insured is sometimes modeled with a negative binomial distribution.

Please note that some textbooks define the negative binomial random variable X as the number of failures (instead of number of trials N) before k-th success. Instead of memorizing a new set of formulas for the probability mass function, mean, variance, and the moment generating function for X, you simply transform X to N by setting $X + K = N$.

To find $f_X(x)$, note that the probability of having x number of failures before k-th success is the same as the probability of having a total of $x+k$ trials before k-th success.

$$f_X(x) = f_N(x+k) = C_{x+k-1}^{k-1} p^k (1-p)^x$$

Or $f_X(x) = f_N(x+k) = \dfrac{k}{x+k} \times$ [binomial probability of k successes in $x+k$ trials]

$$= \frac{k}{x+k} \times C_{x+k}^{k} p^k (1-p)^x$$

$$E(X) = E(N = X+k) - k = \frac{k}{p} - k = k\left(\frac{1}{p} - 1\right)$$

$$Var(X) = Var(N = X+k) = k\left(\frac{1}{p^2} - \frac{1}{p}\right)$$

To find the moment generating function of X, we use the formula $M_{aX+b}(t) = M_{aX}(t)e^{bt}$. Please note that $X = N - k$.

$$\Rightarrow \qquad M_X(t) = M_{N-k}(t) = M_N(t)e^{-kt} = \left[\frac{pe^t}{1-(1-p)e^t}\right]^k e^{-kt} = \left[\frac{p}{1-(1-p)e^t}\right]^k$$

Sample Problems and Solutions

Problem 1

You roll a fair die repeatedly. Let N = the number of trials needed to roll four 5's. Find $E(N)$ and $Var(N)$.

Solution

N is a negative binomial random variable with parameter $k = 4$ and $p = \dfrac{1}{4}$

$$E(N) = \frac{k}{p} = \frac{4}{\frac{1}{6}} = 24$$

$E(N)$ formula is intuitive. On average, you need to roll the die 6 times to get one 5. To get four 5's, you need to roll 4*6=24 times.

$$Var(N) = k\left(\frac{1}{p^2} - \frac{1}{p}\right) = 4\left(6^2 - 6\right) = 4(30) = 120$$

The formula for $Var(N)$ is not intuitive. You have to memorize it.

Problem 2

A negative binomial distribution X =0,1,2,…has parameters k and p. If p is cut in half, $Var(X)$ will increase by 500%. Calculate $\dfrac{E(X)}{Var(X)}$ before p is cut in half.

Solution

Because X =0,1,2,…, X = # of failures before k-th success.

$$E(X) = k\left(\frac{1}{p} - 1\right), \quad Var(X) = k\left(\frac{1}{p^2} - \frac{1}{p}\right) \quad \Rightarrow \quad \frac{E(X)}{Var(X)} = \frac{k\left(\frac{1}{p} - 1\right)}{k\left(\frac{1}{p^2} - \frac{1}{p}\right)} = p$$

To find p, we need to use the information that "If p is cut in half, $Var(X)$ will increase by 500%."

$Var(X)$ if p is cut in half : $\quad Var(X) = k\left[\dfrac{1}{(p/2)^2} - \dfrac{1}{p/2}\right]$

$Var(X)$ prior to p is cut in half: $\quad Var(X) = k\left(\dfrac{1}{p^2} - \dfrac{1}{p}\right)$

$$\Rightarrow \quad k\left[\frac{1}{(p/2)^2} - \frac{1}{p/2}\right] = (1 + 500\%)k\left(\frac{1}{p^2} - \frac{1}{p}\right), \quad p = \frac{1}{2}$$

Prior to p being cut in half, $\quad E(X)/Var(X) = p = 1/2$

Problem 3

A fast food restaurant with a huge daily customer base is selling an experimental (and more expensive) low-carb meal choice among three well-established cheaper meal choices. The profit for the new low-carb meal choice is $2 per meal sold; the profit for any of the three well-established meal choices is $1 per meal sold.

Not sure how many customers will order the low-carb meal choice, the restaurant manager decides to set the food supply (which is more expensive) for the low-carb meal to the level that a maximum of 25 low-carb meals can be served on a given day when the low-carb meal is on trial.

Each visiting customer chooses one and only one meal option among four options (the experimental low-carb meal plus three well-established meal options). Each customer orders his or her meal independent of the order by any other customer. Customers are three times as likely to order any of the three existing meal choices as to order the experimental low-carb meal choice.

Suppose the restaurant closes its doors immediately after the 25 low-carb meals are sold out. Calculate the expected profit the restaurant will make on all the meals sold on a given day when the low-carb meal is on trial.

Solution

The problem is set up in such a way that a negative binomial distribution can be used to model the number of visiting customers before the 25$^{\text{th}}$ low-carb meal is sold. Notice the wording "A fast restaurant with a huge daily customer base …" A huge daily customer base reflects the fact that the negative binomial random variable can be very large (can be $+\infty$ theoretically). If the customer base is very small, then the negative binomial distribution is not appropriate to model the number of visiting customers before the 25$^{\text{th}}$ low-carb meal is sold.

Let

N = # of visiting customers before the 25$^{\text{th}}$ low-carb meal is sold
X = total profit made on all the meals sold while the restaurant is still open.

$$X = 25(2) + (N - 25)(1)$$

The above equation says that for the 25 customers who have bought the new low-carb meal, the restaurant makes \$2 profit per customer; for the remaining $(N - 25)$ customers, the restaurant makes \$1 profit per customer.

Taking the expectation on the above equation, we have

$$E(X) = E\left[25(2) + (N-25)(1)\right] = E(N) + 25$$

We know that $E(N) = \dfrac{k}{p}$.

To find p, the probability that a visiting customer orders a low-carb meal, we use the information that customers are three times as likely to order any of the three conventional meal choices as to order the new meal choice.

$$p + 3(3p) = 1 \quad \Rightarrow \quad p = 0.1, \quad E(N) = \frac{25}{0.1} = 250$$
$$E(X) = E(N) + 25 = 250 + 25 = 275$$

So the restaurant can expect to earn $275 total profit by the time that the 25 low-carb meals are sold out.

Problem 4

A fair die is repeatedly thrown *n* times until the second 5 appears. What is the probability that n =20? What is the probability of having to throw the die more than 20 times (i.e. $n \geq 21$)?

Solution

$$f^{NB}(n=20, k=2, p=1/6) = \frac{2}{20} f^{B}(n=20, k=2, p=1/6)$$
$$= \frac{2}{20} C_{20}^{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{18} = 1.982\%$$

To calculate the probability of throwing the die at least 21 times to get the second 5, we sum up the probability for n =21, 22,…, $+\infty$. The sum should converge as $n \to +\infty$. However, this approach is time-consuming.

The alternate approach is to calculate the probability for *n*=1,2,…,20.

$$P(n=21) = 1 - \left[P(n=1) + P(n=2) + P(n=3) + ... + +P(n=20)\right]$$

This is a lot of work too.

As a shortcut, you can reason that the probability of having to throw the die at least 21 times to get the second 5 is the same as the probability of having a zero or one 5 in 20 trials. Obviously, if you throw the die 20 times but get zero or only one 5, you just have to keep throwing the die more times in order to get a total of two 5's. Having a zero or one 5 in 20 trials is a binomial distribution.

P(throwing the die at least 20 times to get the second 5)
=*P*(having no 5's in 20 trials} + *P*(having one 5 in 20 trials)
$=C_{20}^{0}\left(1/6\right)^{0}\left(5/6\right)^{20}+C_{20}^{1}\left(1/6\right)^{1}\left(5/6\right)^{19}$ =0.1304

Problem 5

The total number of claims in a year incurred by a small block of auto insurance policies is modeled as a negative binomial distribution with p=0.2 and k=5. If there is a claim, the claim amount is $2,000 per claim for all claims. Find the probability that the total amount of annual claims incurred by this block of auto insurance will exceed $100,000.

Solution

Let

N = total # of claims in a year by the group of the insured.
X = total dollar amount of the annual claims by the group of the insured

Then $X=2,000N$.
$P\left(X>100,000\right)=P\left(2,000N>100,000\right)=P\left(N>50\right)$

To find $P\left(N>50\right)$, be careful on how you interpret N, the negative binomial random variable. Because N is the total number of claims in year, theoretically N can be zero. Thus, you should interpret N as the number of failures, not the number of trials, before the 5th success. If you interpret N as the number of trials, then N =5,6,7,…,+∞. You essentially set 5 as the minimum number of claims. This is clearly inappropriate because the problem doesn't say that the annual number of claims must be at least 5.

Next, you convert N to the number of trials before the 5th success:

# of failures + k = # of trials

$\Rightarrow$ more than 50 failures = more than 55 trials.

f^{NB}(more than 50 failures before 5th success, p =0.2)
= f^{NB}(more than 55 trials before 5th success, p =0.2)

= f^B (more than 5 successes in 55 trials, p =0.2)
= f^B (0 success in 55 trials, p =0.2) + f^B (1 success in 55 trials, p =0.2)
+ f^B (2 success in 55 trials, p =0.2) + f^B (3 success in 55 trials, p =0.2)
+ f^B (4 success in 55 trials, p =0.2)

In other words, to have more than 55 trials before getting the 5th success, you need to have at most 4 successes in the 55 trials. The probability of having at most 4 successes in 55 trials is a binomial probability.

f^B (more than 5 successes in 55 trials, p =0.2)

$$=\sum_{x=0}^{4} C_{55}^{x} 0.2^{x} 0.8^{55-x}$$

$$=C_{55}^{0} 0.8^{55}+C_{55}^{1} 0.2\left(0.8^{54}\right)+C_{55}^{2}\left(0.2^{2}\right)\left(0.8^{53}\right)+C_{55}^{3}\left(0.2^{3}\right)\left(0.8^{52}\right)+C_{55}^{4}\left(0.2^{4}\right)\left(0.8^{51}\right)$$

= 0.865%

Tip: how to check your result for complex calculations

I often use Microsoft Excel to check complex calculations. For example, in Excel the formula for calculating $\sum_{x=0}^{4} C_{55}^{x} 0.2^{x} 0.8^{55-x}$ is

BINOMDIST(# of successes, # of trials, p, indicator)
=BINOMDIST(k, n, p, indicator)= BINOMDIST (4, 55, 0.2, True)

If you set the indicator = True, you'll calculate the cumulative density, the probability of having **at most** k successes in n trials, $\sum_{x=0}^{k} C_{n}^{x} p^{x}(1-p)^{n-x}$.

If you set the indicator = False, you'll calculate the probability of having **exactly** k successes in n trials, $C_{n}^{k} p^{k}(1-p)^{n-k}$.

Excel also has formulas for negative binomial, Poisson, exponential densities. You can use Excel "Help" menu to learn how the formulas.

Problem 6

At 4:00 pm on a Friday afternoon, an overworked claim adjustor realized that he still had five pending claims to review. To accelerate his work, the claim adjustor decided to throw a die to determine the dollar amount to be awarded for each claim. He kept throwing a die until he got the second 1 or second 6 or until he had thrown the die 5 times, at which time he stopped throwing the die. At each throw of the die, he picked a claim file randomly and awarded $10,000 to that claim, regardless of the actual dollar amount that claim deserved. Any claims left after the last throw of the die were awarded $5,000 per claim, regardless of the actual circumstance of the claims.

Find
1. the expected total dollar amount awarded to the five claims.
2. the variance of the total dollar amount awarded to the five claims.

Solution

Let N =# of times the dice was thrown before the second 1 or second 6. Then N has a negative binomial distribution with k =2 and $p=2/6=1/3$.

Let X = total dollar amount awarded to the 5 claims.

Let's come up with a table listing all possible values of $X(n)$ and $f(n)$. Then we can calculate mean and variance of X using the standard formula:

$$E(X)=\sum_{n=2}^{+\infty}X(n)f(n), \qquad E(X^2)=\sum_{n=2}^{+\infty}X^2(n)f(n),$$

$$Var(X)=E(X^2)-E^2(X)$$

$N=n$	$P(N=n)=f^{NB}(n,k=2,p=1/3)$ $=\frac{2}{n}f^{B}(n,k=2,p=1/3)=\frac{2}{n}\times\frac{n!}{2!(n-2)!}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{n-2}$	X (K=$1,000)
$n=2$	$\frac{2}{2}\times\frac{2!}{2!(2-2)!}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{2-2}=\left(\frac{1}{3}\right)^2=0.1111$	10(2)+5(3)=35K
$n=3$	$\frac{2}{3}\times\frac{3!}{2!(3-2)!}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{3-2}=0.1481$	10(3)+5(2)=40K
$n=4$	$\frac{2}{4}\times\frac{4!}{2!(4-2)!}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{4-2}=0.1481$	10(4)+5(1)=45K
$n\geq 5$	$1-\left[P(N=2)+P(N=3)+P(N=4)\right]$ =1- (0.1111+0.1481+0.1481)=0.5927	10(5)=50K

To use BA II Plus 1-V Statistics Worksheet, we'll scale up the probabilities:

X	$P(X=x)$	$1000\,P(X=x)$
35	.1111	1111
45	.1481	1481
55	.1481	1481
65	.5927	5927

Plugging the numbers in BA II Plus, we get:

$E(X)$=57.224(K)=$57,224, σ_X =10.65710205(K)

$Var(X)=113.573824\left(K^2\right)=113.573824\left(\$^2 1000,000\right)=113,573,824\left(\$^2\right)$

Please note the unit for $Var(X)$ is dollar squared ($\$^2$).

We write the values of X as 35K,40K,45K,and 50K instead of $35,000,$40,000,$45,000, and $50,000. This significantly reduces the number of times we have to press calculator keys, hence reducing possible errors of accidentally dropping or adding a zero.

Homework for you: #11 Nov 2001.

Chapter 16 Hypergeometric distribution

Imagine a shipment has seven defective parts and ten good parts. You want to find out if you randomly take out five parts, how likely is it to get three defective parts and two good parts.

Your first reaction might be to use binomial distribution with parameters $n=5$ and $p=7/17$.

Pr(having 3 defective parts out of 5 parts)= $C_5^3\left(7/17\right)^3\left(10/17\right)^2=0.2416$

After close examination, you will find that binomial distribution is not a good fit here. If you use binominal distribution, essentially you assume (1) you have 5 independent draws, and (2) the chance of a defective part being chosen at each of the 5 draws is 7/17. The only way to meet these two conditions is that after you randomly choose your first part, whether it is defective or good, you put it back before your second random draw. Then after you randomly take out the second part, you put the second part back before making the third draw. And so on.

If you do not put the first part back, your chance of getting a defective part in the second draw is 4/16 (if you get a defective part in the first draw) or 5/16 (if you get a good part in the first draw). Now the outcome of a previous trial affects the outcome of the next trial and the probability of having a success in different trials is different now.

You can see that the only way to meet the conditions for binomial distribution is "putting things back" before making the next draw. In mathematical terms, "putting things back" before making the next draw is sampling with replacement. Binomial distribution is sampling with replacement.

In this problem, however, we don't put things back (so here we have sampling without replacement). We permanently take out the first part before making our second draw. We permanently take out the second part before making our third draw and so on.

To find the probability that exactly 3 out of the 5 randomly chosen parts are defective, we reason that out of the total of C_{17}^5 ways of getting 5 parts, we have C_7^3 ways of getting 3 defective parts and C_{10}^2 ways of getting 2 good parts. Thus, the probability that exactly 3 out of the 5 randomly chosen parts are defective is:

$$\frac{C_7^3\times C_{10}^2}{C_{17}^5}$$

The general formula for a hypergeometric distribution:

A finite population has N objects, of which K objects are special and N-K objects are ordinary. If m objects are randomly chosen from the population, the probability that out of the m objects chosen, x objects are special is:

$$f(x) = \frac{C_K^x \times C_{N-K}^{m-x}}{C_N^m}$$

When you apply the above formula to a problem, please do two quality checks:

First, make sure the combinatorial elements in the nominator add up to the combinatorial elements in the denominator. Notice in the formula above, we have

$$x + (m - x) = m, \quad K + (N - K) = N$$

You want to make sure that these two equations are satisfied. Otherwise, you did something wrong.

Second, make sure you don't have any negative factorials. In C_K^x, make sure $x \le K$; in C_{N-K}^{m-x}, make sure $m - x \le N - K$; in C_N^m, make sure $m \le N$. If you get a negative factorial, you did something wrong.

Please note that if both N and K are big, the hypergeometric distribution and binomial distribution give roughly the same result. So if N and K are big, we do not have to worry about whether we sample with replacement or without replacement. We can use just use binomial distribution to approximate the hypergeometric distribution by setting $p = K/N$.

Sample Problems and Solutions

Problem 1

Among 200 people working for a large actuary department of an insurance company, 120 are actuaries and 80 are support staff. If 8 people are randomly chosen to attend a brainstorm meeting with company executives, what is the probability that exactly 5 actuaries are chosen to attend the meeting?

Solution

$$\frac{C_{120}^5 \times C_{80}^3}{C_{200}^8} = 0.2842$$

Two quality checks: 5+3=8, 120+80=120. In addition, we don't have any negative factorials.

If we use binomial distribution to approximate, we have $p = 120/200 = 0.6$.
$C_8^5 0.6^5 0.4^3$ =0.2787

You see that if both N and K are big, binomial distribution gives good approximation to hypergeometric distribution.

I think this is all you need to know about hypergeometric distribution. Don't bother memorizing the mean and variance formulas for hypergeometric distribution. You are better off spending your time on something else.

Homework for you: rework the problem in this chapter.

Chapter 17 Uniform distribution

Key formulas

If the random variable X is uniformly distributed over $[a,b]$ where $b>a$, then

$$f(x)=\frac{1}{b-a},\quad E(X)=\frac{a+b}{2},\quad \sigma_X=\frac{b-a}{2\sqrt{3}}$$

Proof.

$$E(X)=\int_a^b x\frac{1}{b-a}dx=\frac{1}{b-a}\left[\frac{1}{2}x^2\right]_a^b=\frac{1}{b-a}\frac{1}{2}\left(b^2-a^2\right)=\frac{b+a}{2}$$

$$E\left(X^2\right)=\int_a^b x^2\frac{1}{b-a}dx=\frac{1}{b-a}\left[\frac{1}{3}x^3\right]_a^b=\frac{1}{b-a}\frac{1}{3}\left(b^3-a^3\right)=\frac{b^2+ab+a^2}{3}$$

$$Var(X)=E\left(X^2\right)-E^2(X)=\frac{b^2+ab+a^2}{3}-\left(\frac{b+a}{2}\right)^2=\frac{(b-a)^2}{12}$$

$$\sigma_X=\sqrt{Var(X)}=\frac{b-a}{2\sqrt{3}}$$

Sample Problems and Solutions

Problem 1

Loss for an auto insurance policy is uniformly distributed over $[0,L]$. If the deductible is 500, then the expected claim payment with the deductible is only $1/9$ of the expected claim payment without the deductible. Find the expected claim payment without the deductible.

Solution

Let X represents the loss incurred by the auto insurance policy. X is uniformly distributed over $[0,L]$. We have

$$f(x)=\frac{1}{L},\qquad E(X)=\frac{L}{2}$$

Let Y represent the claim payment with a deductible of 500. Y is distributed as follows:

$$Y = \begin{cases} 0 & \text{if } X \le 500 \\ X - 500 & \text{if } X > 500 \end{cases}$$

$$E(Y) = \int_0^L y(x) f(x) dx = \int_0^{500} 0 f(x) dx + \int_{500}^L (x-500) f(x) dx = \int_{500}^L (x-500) f(x) dx$$

$$= \frac{1}{2L}\left[(x-500)^2\right]_{500}^L = \frac{1}{2L}(L-500)^2$$

If there is no deductible, then the expected claim will be $E(Y) = E(X) = \dfrac{L}{2}$.

Because the expected claim with the deductible is 1/9 of the expected claim with no deductible, we have

$$\frac{1}{2L}(L-500)^2 = \frac{1}{9}\left(\frac{L}{2}\right), \quad \Rightarrow \quad (L-500)^2 = \left(\frac{L}{3}\right)^2$$

Please note that L must exceed the deductible 500. So,

$$L - 500 = \frac{L}{3}, \quad \Rightarrow \quad L = 750$$

The expected claim without a deductible is $L/2$ =375.

Problem 2

In a policy year, a low risk auto insurance policyholder can incur either one loss with probability of p or no loss with probability of 1- p. p is uniformly distributed over [0, 0.3]. If there is a loss, loss is uniformly distributed over [500, 2,500].

Assume that the probability of having a loss and the loss amount are independent. Find the expected loss incurred by the policyholder in a policy year.

Solution

Let N = # of losses in a year, X = amount of individual losses in a year, Y is the total loss amount in a year.

$Y = N X$, $\quad \Rightarrow \quad E(Y) = E(N)E(X)$ (because N and X are independent)

$E(X)=\dfrac{500+2,500}{2}=1,500$ because X is uniform over [500, 2,500]

To find the mean $E(N)$, we'll use the double expectation theorem (to be explained in a future chapter):

$$E(N)=E_P\left[E(N|p)\right]=\int_0^{0.3}E(N|p)\,f(p)\,dp$$

The above equation says that to find $E(N)$, we first calculate $E(N|p)$ assuming that p is a known constant. This gives us $E(N|p)$, the conditional mean given a known constant p. $E(N|p)$ can be interpreted as the contribution to $E(N)$ made by a known constant p. Next, we calculate all p's contribution to $E(N)$ by integrating $E(N|p)\,f(p)$ over $p\in[0,0.3]$.

Given p (i.e. if we fix p), N is a Bernoulli random variable with parameter p.

$\Rightarrow \quad E(N|p)=p$

$\Rightarrow \quad E(N)=E_P\left[E(N|p)\right]=\int_0^{0.3}E(N|p)\,f(p)\,dp=\int_0^{0.3}pf(p)\,dp=E(P)=\dfrac{0.3}{2}=0.15$

Finally, $E(Y)$=0.15*1,500=225

Alternatively,

Average loss amount
= average probability of having a claim × average claim amount if there is a loss

average probability of having a loss = 0.3/2 = 0.15
average loss amount if there's a loss = (500+2,500)/2 =1,500

Average loss amount = 0.15(1,500) = 225.

Homework for you: #38 May 2000; #29 Nov 2001.

Chapter 18 Exponential distribution

Key Points

Gain a deeper understanding of exponential distribution:

Why does exponential distribution model the time elapsed before the first or next random event occurs?

Exponential distribution lacks memory. What does this mean?

Understand and use the following integration shortcuts:

For any $\theta > 0$ and $a \geq 0$:

$$\int_a^{+\infty} \frac{1}{\theta} e^{-x/\theta} dx = e^{-a/\theta}$$

$$\int_a^{+\infty} x\left(\frac{1}{\theta} e^{-x/\theta}\right) dx = (a+\theta) e^{-a/\theta}$$

$$\int_a^{+\infty} x^2\left(\frac{1}{\theta} e^{-x/\theta}\right) dx = \left[(a+\theta)^2 + \theta^2\right] e^{-a/\theta}$$

You will need to understand and memorize these shortcuts to quickly solve integrations in the heat of the exam. Do not attempt to do integration by parts during the exam.

Explanations

Exponential distribution is the continuous version of geometric distribution. While geometric distribution describes the probability of having N trials before the first or next success (success is a random event), exponential distribution describes the probability of having time T elapse before the first or next success.

Let's use a simple example to derive the probability density function of exponential distribution. Claims for many insurance policies seem to occur randomly. Assume that on average, claims for a certain type of insurance policy occur once every 3 months. We want to find out the probability that T months will elapse before the next claim occurs.

To find the pdf of exponential distribution, we take advantage of geometric distribution, whose probability mass function we already know. We divide each month into n intervals, each interval being one minute. Since there are 30*24*60=43,200 minutes in a month (assuming there are 30 days in a month), we convert each month into 43,200 intervals. Because on average one claim occurs every 3 months, we assume that the chance of a claim occurring in one minute is

$$\frac{1}{3\times 43,200}$$

How many minutes must elapse before the next claim occurs? We can think of one minute as one trial. Then the probability of having *m* trials (i.e. *m* minutes) before the first success is a geometric distribution with

$$p=\frac{1}{3\times 43,200}$$

Instead of finding the probability that it takes **exactly** *m* minutes to have the first claim, we'll find the probability that it takes *m* **minutes or less** to have the first claim. The latter is the cumulative distribution function which is more useful.

P (it takes *m* minutes or less to have the first claim)
=1 – *P* (it takes more than *m* minutes to have the first claim)

The probability that it takes more than *m* trials before the first claim is $(1-p)^m$. To see why, you can reason that to have the first success only after m trials, the first *m* trials must all end with failures. The probability of having *m* failures in *m* trials is $(1-p)^m$.

Therefore, the probability that it takes *m* trials or less before the first success is $1-(1-p)^m$.

Now we are ready to find the pdf of T:

$P(T\leq t)=P$ (43,200t trials or fewer before the 1st success)

$$=1-\left(1-\frac{1}{3\times 43,200}\right)^{43,200t}=1-\left[\left(1-\frac{1}{3\times 43,200}\right)^{-3\times 43,200t}\right]^{-t/3}\approx 1-e^{-t/3}$$

Of course, we do not need to limit ourselves by dividing one month into intervals of one minute. We can divide, for example, one month into *n* intervals, with each interval of 1/1,000,000 of a minute. Essentially, we want $n\to+\infty$.

$P(T\leq t)=P$ (nt trials or fewer before the 1st success)

$$=1-\left(1-\frac{1}{3n}\right)^{nt}=1-\left[\left(1-\frac{1}{3n}\right)^{-3n}\right]^{-t/3}=1-e^{-t/3}\quad (\text{as } n\to+\infty)$$

If you understand the above, you should have no trouble understanding why exponential distribution is often used to model time elapsed until the next random event happens.

Here are some examples where exponential distribution can be used:

- Time until the next claim arrives in the claims office.
- Time until you have your next car accident.
- Time until the next customer arrives at a supermarket.
- Time until the next phone call arrives at the switchboard.

General formula:

Let T=time elapsed (in years, months, days, etc.) before the next random event occurs.

$$F(t)=P(T\le t)=1-e^{-t/\theta},\quad f(t)=\frac{1}{\theta}e^{-t/\theta}, \text{ where } \theta=E(T)$$
$$P(T>t)=1-F(t)=e^{-t/\theta}$$

Alternatively,

$$F(t)=P(T\le t)=1-e^{-\lambda\theta},\quad f(t)=\lambda e^{-\lambda\theta}, \text{ where } \lambda=\frac{1}{E(T)}$$
$$P(T>t)=1-F(t)=e^{-\lambda\theta}$$

Mean and variance: $E(T)=\theta=\dfrac{1}{\lambda}$, $Var(T)=\theta^2=\dfrac{1}{\lambda^2}$

Like geometric distribution, exponential distribution lacks memory:

$$P(T>a+b\,|\,T>a)=P(T>b)$$

We can easily derive this:

$$P(T>a+b\,|\,T>a)=\frac{P(T>a+b\cap T>a)}{P(T>a)}=\frac{P(T>a+b)}{P(T>a)}=\frac{e^{-(a+b)/\theta}}{e^{-a/\theta}}=e^{-b/\theta}=P(T>b)$$

In plain English, this lack of memory means that if a component's time to failure follows exponential distribution, then the component does not remember how long it has been working (i.e. does not remember wear and tear). At any moment when it is working, the component starts fresh as if it were completely new. At any moment while the component is working, if you reset the clock to zero and count the time elapsed from then until the component breaks down, the time elapsed before a breakdown is always exponentially distributed with the identical mean.

This is clearly an idealized situation, for in real life wear and tear does reduce the longevity of a component. However, in many real world situations, exponential distribution can be used to approximate the actual distribution of time until failure and still give a reasonably accurate result.

A simple way to see why a component can, at least by theory, forget how long it has worked so far is to think about geometric distribution, the discrete counterpart of exponential distribution. For example, in tossing a coin, you can clearly see why a coin doesn't remember its past success history. Since getting heads or tails is a purely random event, how many times you have tossed a coin so far before getting heads really should NOT have any bearing on how many more times you need to toss the coin to get heads the second time.

The calculation shortcuts are explained in the following sample problems.

Sample Problems and Solutions

Problem 1

The lifetime of a light bulb follows exponential distribution with a mean of 100 days. Find the probability that the light bulb's life …

(1) Exceeds 100 days
(2) Exceeds 400 days
(3) Exceeds 400 days given it exceeds 100 days.

Solution

Let T = # of days before the light bulb burns out.

$F(t) = 1 - e^{-t/\theta}$, where $\theta = E(T) = 100$
$P(T > t) = 1 - F(t) = e^{-t/\theta}$

$P(T > 100) = 1 - F(100) = e^{-100/100} = e^{-1} = 0.3679$
$P(T > 400) = 1 - F(400) = e^{-400/100} = e^{-4} = 0.0183$

$$P(T > 400 | T > 100) = \frac{P(T > 400)}{P(T > 100)} = \frac{e^{-400/100}}{e^{-100/100}} = e^{-3} = 0.0498$$

Or use the lack of memory property of exponential distribution:

$$P(T > 400 | T > 100) = P(T > 400 - 100) = e^{-300/100} = e^{-3} = 0.0498$$

Problem 2

The length of telephone conversations follows exponential distribution. If the average telephone conversation is 2.5 minutes, what is the probability that a telephone conversation lasts between 3 minutes and 5 minutes?

Solution

$$F(t) = 1 - e^{-t/2.5}$$
$$P(3 < T < 5) = \left(1 - e^{-5/2.5}\right) - \left(1 - e^{-3/2.5}\right) = e^{-3/2.5} - e^{-5/2.5} = 0.1659$$

Problem 3

The random variable T has an exponential distribution with pdf $f(t) = \frac{1}{2} e^{-t/2}$.

Find $E(T | T > 3)$, $Var(T | T > 3)$, $E(T | T \le 3)$, $Var(T | T \le 3)$.

Solution

First, let's understand the conceptual thinking behind the symbol $E(T | T > 3)$. Here we are only interested in $T > 3$. So we reduce the original sample space $T \in [0, +\infty]$ to $T \in [3, +\infty]$. The pdf in the original sample space $T \in [0, +\infty]$ is $f(t)$; the pdf in the reduced sample space $t \in [3, +\infty]$ is $\frac{f(t)}{P(T > 3)}$. Here the factor $\frac{1}{P(T > 3)}$ is a normalizing constant to make the total probability in the reduced sample space add up to one:

$$\int_3^{+\infty} \frac{f(t)}{P(T > 3)} dt = \frac{1}{P(T > 3)} \int_3^{+\infty} f(t) dt = \frac{1}{P(T > 3)} \times P(T > 3) = 1$$

Next, we need to calculate $E(T|T>3)$, the expected value of T in the reduced sample space $T\in[3,+\infty]$:

$$E(T|T>3)=\int_{\substack{\text{Reduced}\\ \text{Sample space}\\ T\in[3,+\infty]}} t\frac{f(t)}{P(T>3)}dt=\frac{1}{P(T>3)}\int_3^{+\infty} t\,f(t)dt=\frac{1}{1-F(3)}\int_3^{+\infty} t\,f(t)dt$$

$$1-F(3)=e^{-3/2}$$

$$\int_3^{+\infty} tf(t)dt=5e^{-3/2} \quad \text{(integration by parts)}$$

$$E(T|T>3)=\frac{1}{1-F(3)}\int_3^{+\infty} tf(t)dt=\frac{5e^{-3/2}}{e^{-3/2}}=5$$

Here is another approach. Because T does not remember wear and tear and always starts anew at any working moment, the time elapsed since T=3 until the next random event (i.e. $T-3$) has exponential distribution with an identical mean of 2. In other words, $(T-3|T>3)$ is exponentially distributed with an identical mean of 2.

So $E(T-3|T>3)=2$.

$$E(T|T>3)=E(T-3|T>3)+3=2+3=5$$

Next, we will find $Var(T|T>3)$.

$$E(T^2|T>3)=\frac{1}{\Pr(T>3)}\int_3^{+\infty} t^2 f(t)dt=\frac{1}{\Pr(T>3)}\int_3^{+\infty} t^2\frac{1}{2}e^{-t/2}dt$$

$$\int_3^{+\infty} t^2\frac{1}{2}e^{-t/2}dt=29e^{-3/2} \quad \text{(integration by parts)}$$

$$E(T^2|T>3)=\frac{29e^{-3/2}}{e^{-3/2}}=29$$

$$Var(T|T>3)=E(T^2|T>3)-E^2(T|T>3)=29-5^2=4=\theta^2$$

It is no coincidence that $Var(T|T>3)$ is the same as $Var(T)$. To see why, we know $Var(T|T>3) = Var(T-3|T>3)$. This is because $Var(X+c) = Var(X)$ stands for any constant c.

Since $(T-3|T>3)$ is exponentially distributed with an identical mean of 2, then

$$Var(T-3|T>3) = \theta^2 = 2^2 = 4.$$

Next, we need to find $E(T|T \le 3)$.

$$E(T|T<3) = \int_0^3 t\frac{f(t)}{\Pr(T<3)}dt = \frac{1}{F(3)}\int_0^3 tf(t)dt$$

$$F(3) = 1 - e^{-3/2}$$

$$\int_0^3 tf(t)dt = \int_0^{+\infty} tf(t)dt - \int_3^{+\infty} tf(t)dt$$

$$\int_0^{+\infty} tf(t)dt = E(T) = 2$$

$$\int_3^{+\infty} tf(t)dt = 5e^{-3/2} \text{ (we already calculated this)}$$

$$E(T|T<3) = \frac{1}{F(3)}\int_0^3 tf(t)dt = \frac{2-5e^{-3/2}}{1-e^{-3/2}}$$

Here is another way to find $E(T|T<3)$.

$$E(T) = E(T|T<3)\times P(T<3) + E(T|T>3)\times P(T>3)$$

The above equation says that if we break down T into two groups, T >3 and T <3, then the overall mean of these two groups as a whole is equal to the weighted average mean of these groups.

Also note that $P(T=3)$ is not included in the right-hand side because the probability of a continuous random variable at any single point is zero. This is similar to the concept that the mass of a single point is zero.

Of course, you can also write:

$$E(T) = E(T|T \le 3) \times P(T \le 3) + E(T|T > 3) \times P(T > 3)$$

Or $E(T) = E(T|T < 3) \times P(T < 3) + E(T|T \ge 3) \times P(T \ge 3)$

You should get the same result no matter which formula you use.

$$E(T) = E(T|T < 3) \times P(T < 3) + E(T|T > 3) \times P(T > 3)$$

$$\Rightarrow \quad E(T|T < 3) = \frac{E(T) - E(T|T > 3) \times P(T > 3)}{P(T < 3)}$$

$$\Rightarrow \quad E(T|T < 3) = \frac{\theta - (\theta + 3)e^{-3/2}}{1 - e^{-3/2}} = \frac{2 - 5e^{-3/2}}{1 - e^{-3/2}}$$

Next, we will find $E(T^2|T < 3)$:

$$E(T^2|T < 3) = \frac{1}{P(T < 3)} \int_0^3 t^2 f(t) dt = \frac{1}{P(T < 3)} \int_0^3 t^2 \frac{1}{2} e^{-t/2} dt$$

$$\int_0^3 t^2 \frac{1}{2} e^{-t/2} dt = -\left[(t+2)^2 + 4\right] e^{-x/2} \Big|_0^3 = 8 - 29e^{-3/2}$$

$$E(T^2|T < 3) = \frac{1}{P(T > 3)} \int_0^3 t^2 \frac{1}{2} e^{-t/2} dt = \frac{8 - 29e^{-3/2}}{1 - e^{-3/2}}$$

Alternatively,

$$E(T^2) = E(T^2|T < 3) \times P(T < 3) + E(T^2|T > 3) \times P(T > 3)$$

$$\Rightarrow \quad E(T^2|T < 3) = \frac{E(T^2) - E(T^2|T > 3) \times P(T > 3)}{P(T < 3)}$$

$$= \frac{2\theta^2 - 29 \times P(T > 3)}{P(T < 3)} = \frac{8 - 29e^{-3/2}}{1 - e^{-3/2}}$$

$$Var\left(T\middle|T<3\right)=E\left(T^{2}\middle|T<3\right)-E^{2}\left(T\middle|T<3\right)=\frac{8-29e^{-3/2}}{1-e^{-3/2}}-\left(\frac{2-5e^{-3/2}}{1-e^{-3/2}}\right)^{2}$$

In general, for any exponentially distributed random variable T with mean $\theta>0$ and for any $a\geq 0$:

$T-a\middle|T>a$ is also exponentially distributed with mean θ

$$\Rightarrow \quad E\left(T-a\middle|T>a\right)=\theta,\quad Var\left(T-a\middle|T>a\right)=\theta^{2}$$

$$\Rightarrow \quad E\left(T\middle|T>a\right)=a+\theta,\quad Var\left(T\middle|T>a\right)=\theta^{2}$$

$$E\left(T-a\middle|T>a\right)=\frac{1}{P(T>a)}\int_{a}^{+\infty}(t-a)\,f(t)dt$$

$$\Rightarrow \quad \int_{a}^{+\infty}(t-a)\,f(t)dt=E\left(T-a\middle|T>a\right)\times P(T>a)=\theta e^{-a/\theta}$$

$$E\left(T\middle|T>a\right)=\frac{1}{P(T>a)}\int_{a}^{+\infty}tf(t)dt$$

$$\Rightarrow \quad \int_{a}^{+\infty}tf(t)dt=E\left(T\middle|T>a\right)\times P(T>a)=(\theta+a)e^{-a/\theta}$$

$$E\left(T\middle|T<a\right)=\frac{1}{P(T<a)}\int_{0}^{a}tf(t)dt$$

$$\Rightarrow \quad \int_{0}^{a}tf(t)dt=E\left(T\middle|T<a\right)\times P(T<a)=E\left(T\middle|T<a\right)\left(1-e^{-a/\theta}\right)$$

$$E(T)=E\left(T\middle|T<a\right)\times P(T<a)+E\left(T\middle|T>a\right)\times P(T>a)$$

$$\Rightarrow \quad \theta=E\left(T\middle|T<a\right)\times(1-e^{-a/\theta})+E\left(T\middle|T>a\right)\times e^{-a/\theta}$$

$$E\left(T^{2}\right)=E\left(T^{2}\middle|T<a\right)\times P(T<a)+E\left(T^{2}\middle|T>a\right)\times P(T>a)$$

You do not need to memorize the above formulas. However, make sure you understand the logic behind these formulas.

Before we move on to more sample problems, I will give you some integration-by-parts formulas for you to memorize. These formulas are critical to you when solving exponential distribution-related problems in 3 minutes. You should memorize these formulas to avoid doing integration by parts during the exam.

Formulas you need to memorize:

For any $\theta > 0$ and $a \geq 0$

$$\int_a^{+\infty} \frac{1}{\theta} e^{-x/\theta} dx = e^{-a/\theta} \qquad (1)$$

$$\int_a^{+\infty} x\left(\frac{1}{\theta} e^{-x/\theta}\right) dx = (a+\theta) e^{-a/\theta} \qquad (2)$$

$$\int_a^{+\infty} x^2\left(\frac{1}{\theta} e^{-x/\theta}\right) dx = \left[(a+\theta)^2 + \theta^2\right] e^{-a/\theta} \qquad (3)$$

You can always prove the above formulas using integration by parts. However, let me give an intuitive explanation to help you memorize them.

Let X represent an exponentially random variable with a mean of θ, and $f(x)$ is the probability distribution function, then for any $a \geq 0$, Equation (1) represents $P(X > a) = 1 - F(a)$, where $F(x) = 1 - e^{-x/\theta}$ is the cumulative distribution function of X. You should have no trouble memorizing Equation (1).

For Equation (2), from Sample Problem 3, we know

$$\int_a^{+\infty} xf(x)dx = E(X \mid X > a) \times P(X > a) = (a+\theta) e^{-a/\theta}$$

To understand Equation (3), note that

$$P(X > a) = e^{-a/\theta}$$

$$\int_a^{+\infty} x^2 f(x) dx = E(X^2 \mid X > a) \times P(X > a)$$

$$E(X^2 \mid X > a) = E^2(X \mid X > a) + Var(X \mid X > a)$$

$$E^2(X \mid X > a) = (a+\theta)^2, \quad Var(X \mid X > a) = \theta^2$$

Then

$$\int_a^{+\infty} x^2 f(x) dx = \left[(a+\theta)^2 + \theta^2\right] e^{-a/\theta}$$

We can modify Equation (1),(2),(3) into the following equations:

For any $\theta > 0$ and $b \ge a \ge 0$

$$\int_a^b \frac{1}{\theta} e^{-x/\theta} dx = e^{-a/\theta} - e^{-b/\theta} \quad (4)$$

$$\int_a^b x\left(\frac{1}{\theta} e^{-x/\theta}\right) dx = (a+\theta) e^{-a/\theta} - (b+\theta) e^{-b/\theta} \quad (5)$$

$$\int_a^b x^2\left(\frac{1}{\theta} e^{-x/\theta}\right) dx = \left[(a+\theta)^2 + \theta^2\right] e^{-a/\theta} - \left[(b+\theta)^2 + \theta^2\right] e^{-b/\theta} \quad (6)$$

We can easily prove the above equation. For example, for Equation (5):

$$\int_a^b x\left(\frac{1}{\theta} e^{-x/\theta}\right) dx = \int_a^{+\infty} x\left(\frac{1}{\theta} e^{-x/\theta}\right) dx - \int_b^{+\infty} x\left(\frac{1}{\theta} e^{-x/\theta}\right) dx = (a+\theta) e^{-a/\theta} - (b+\theta) e^{-b/\theta}$$

We can modify Equation (1),(2),(3) into the following equations:

For any $\theta > 0$ and $a \ge 0$

$$\int \frac{1}{\theta} e^{-x/\theta} dx = -e^{-x/\theta} + c \quad (7)$$

$$\int x\left(\frac{1}{\theta} e^{-x/\theta}\right) dx = -(x+\theta) e^{-x/\theta} + c \quad (8)$$

$$\int x^2\left(\frac{1}{\theta} e^{-x/\theta}\right) dx = -\left[(x+\theta)^2 + \theta^2\right] e^{-x/\theta} + c \quad (9)$$

Set $\lambda = \frac{1}{\theta}$. For any $\lambda > 0$ and $a \ge 0$

$$\int \lambda e^{-\lambda x} dx = -e^{-\lambda x} + c \quad (10)$$

$$\int x\left(\lambda e^{-\lambda x}\right) dx = -\left(x + \frac{1}{\lambda}\right) e^{-\lambda x} + c \quad (11)$$

$$\int x^2\left(\lambda e^{-\lambda x}\right) dx = -\left[\left(x + \frac{1}{\lambda}\right)^2 + \frac{1}{\lambda^2}\right] e^{-\lambda x} + c \quad (12)$$

So you have four sets of formulas. Just remember one set (any one is fine). Equations (4),(5),(6) are most useful (because you can directly apply the formulas), but the formulas are long.

If you can memorize any one set, you can avoid doing integration by parts during the exam. You definitely do not want to calculate messy integrations from scratch during the exam.

Now we are ready to tackle more problems.

Problem 4

After an old machine was installed in a factory, Worker John is on call 24-hours a day to repair the machine if it breaks down. If the machine breaks down, John will receive a service call right away, in which case he immediately arrives at the factory and starts repairing the machine.

The machine's time to failure is exponentially distributed with a mean of 3 hours. Let T represent the time elapsed between when the machine was first installed and when John starts repairing the machine.

Find $E(T)$ and $Var(T)$.

Solution

T is exponentially distributed with mean $\theta = 3$. $F(t) = 1 - e^{-t/3}$.
We simply apply the mean and variance formula:

$$E(T) = \theta = 3, \quad Var(T) = \theta^2 = 3^2 = 9$$

Problem 5

After an old machine was installed in a factory, Worker John is on call 24-hours a day to repair the machine if it breaks down. If the machine breaks down, John will receive a service call right away, in which case he immediately arrives at the factory and starts repairing the machine.

The machine was found to be working today at 10:00 a.m..

The machine's time to failure is exponentially distributed with a mean of 3 hours. Let T represent the time elapsed between 10:00 a.m. and when John starts repairing the machine.

Find $E(T)$ and $Var(T)$.

Solution

Exponential distribution lacks memory. At any moment when the machine is working, it forgets its past wear and tear and starts afresh. If we reset the clock at 10:00 and observe T , the time elapsed until a breakdown, T is exponentially distributed with a mean of 3.

$E(T)=\theta=3$, $\qquad Var(T)=\theta^2=3^2=9$

Problem 6

After an old machine was installed in a factory, Worker John is on call 24-hours a day to repair the machine if it breaks down. If the machine breaks down, John will receive a service call right away, in which case he immediately arrives at the factory and starts repairing the machine.

Today, John happens to have an appointment from 10:00 a.m. to 12:00 noon. During the appointment, he won't be able to repair the machine if it breaks down.

The machine was found working today at 10:00 a.m..

The machine's time to failure is exponentially distributed with a mean of 3 hours. Let X represent the time elapsed between 10:00 a.m. today and when John starts repairing the machine.

Find $E(T)$ and $Var(T)$.

Solution

Let T =time elapsed between 10:00 a.m. today and a breakdown. T is exponentially distributed with a mean of 3. $X=\max(2,T)$.

$$X=\begin{cases}2, \text{ if } T\le 2\\ T, \text{ if } T>2\end{cases}$$

You can also write

$$X=\begin{cases}2, \text{ if } T<2\\ T, \text{ if } T\ge 2\end{cases}$$

As said before, it doesn't matter where you include the point T =2 because the probability density function of a continuous variable at any single point is always zero.

Pdf is always $f(t)=\dfrac{1}{3}e^{-t/3}$ no matter $T \le 2$ or $T > 2$.

$$E(X)=\int_0^{+\infty} x(t)\,f(t)\,dt=\int_0^2 2\left(\frac{1}{3}e^{-t/3}\right)dt+\int_2^{+\infty} t\left(\frac{1}{3}e^{-t/3}\right)dt$$

$$\int_0^2 2\left(\frac{1}{3}e^{-t/3}\right)dt=2\left(1-e^{-2/3}\right)$$

$$\int_2^{+\infty} t\left(\frac{1}{3}e^{-t/3}\right)dt=(2+3)e^{-2/3}=5e^{-2/3}$$

$$E(X)=2\left(1-e^{-2/3}\right)+5e^{-2/3}=2+3e^{-2/3}$$

$$E\left(X^2\right)=\int_0^{+\infty} x^2 f(t)\,dt=\int_0^2 2^2\left(\frac{1}{3}e^{-t/3}\right)dt+\int_2^{+\infty} t^2\left(\frac{1}{3}e^{-t/3}\right)dt$$

$$\int_0^2 2^2\left(\frac{1}{3}e^{-t/3}\right)dt=2^2\left(1-e^{-2/3}\right)=4\left(1-e^{-2/3}\right)$$

$$\int_2^{+\infty} t^2\left(\frac{1}{3}e^{-t/3}\right)dt=\left(5^2+3^2\right)e^{-2/3}=34e^{-2/3}$$

$$E\left(X^2\right)=4\left(1-e^{-2/3}\right)+34e^{-2/3}=4+30e^{-2/3}$$

$$Var(X)=E\left(X^2\right)-E^2(X)=4+30e^{-2/3}-\left(2+3e^{-2/3}\right)^2$$

We can quickly check that $E(X)=2+3e^{-2/3}$ is correct:

$$X=\begin{cases}2, & \text{if } T\le 2\\ T, & \text{if } T>2\end{cases} \quad\Rightarrow\quad X-2=\begin{cases}0, & \text{if } T\le 2\\ T-2, & \text{if } T>2\end{cases}$$

$$\begin{aligned}\Rightarrow\quad E(X-2)&=0\times E\left(T\,|\,T<2\right)\times P(T<2)+E\left(T-2\,|\,T>2\right)\times P(T>2)\\ &=E\left(T-2\,|\,T>2\right)\times P(T>2)=3e^{-2/3}\end{aligned}$$

$$\Rightarrow\quad E(X)=E(X-2)+2=2+3e^{-2/3}$$

You can use this approach to find $E\left(X^2\right)$ too, but this approach isn't any quicker than using the integration as we did above

Problem 7

After an old machine was installed in a factory, Worker John is on call 24-hours a day to repair the machine if it breaks down. If the machine breaks down, John will receive a service call right away, in which case he immediately arrives at the factory and starts repairing the machine.

Today is John's last day of work because he got an offer from another company, but he'll continue his current job of repairing the machine until 12:00 noon if there's a breakdown. However, if the machine does not break by noon 12:00, John will have a final check of the machine at 12:00. After 12:00 noon John will permanently leave his current job and take a new job at another company.

The machine was found working today at 10:00 a.m..

The machine's time to failure is exponentially distributed with a mean of 3 hours. Let X represent the time elapsed between 10:00 a.m. today and John's visit to the machine.

Find $E(X)$ and $Var(X)$.

Solution

Let T =time elapsed between 10:00 a.m. today and a breakdown. T is exponentially distributed with a mean of 3. $X = \min(2, T)$.

$$X = \begin{cases} t, & \text{if } T \le 2 \\ 2, & \text{if } T > 2 \end{cases}$$

Pdf is always $f(t) = \frac{1}{3}e^{-t/3}$ no matter $T \le 2$ or $T > 2$.

$$E(X) = \int_0^2 t\frac{1}{3}e^{-t/3}dt + \int_2^{+\infty} 2\left(\frac{1}{3}e^{-t/3}\right) dt$$

$$\int_0^2 t\frac{1}{3}e^{-t/3}dt = 3 - (2+3)e^{-2/3}$$

$$\int_2^{+\infty} 2\left(\frac{1}{3}e^{-t/3}\right) dt = 2e^{-2/3}$$

$$E(X) = 3 - 5e^{-2/3} + 2e^{-2/3} = 3 - 3e^{-2/3}$$

To find $Var(X)$, we need to calculate $E(X^2)$.

$$E(X^2)=\int_0^{+\infty} x^2(t)f(t)dt=\int_0^2 t^2 f(t)dt+\int_2^{+\infty} 2^2 f(t)dt$$

$$\int_0^2 t^2 f(t)dt=\left[(0+3)^2+3^2\right]e^{-0/3}-\left[(2+3)^2+3^2\right]e^{-2/3}=18-34e^{-2/3}$$

$$\int_2^{+\infty} 2^2 f(t)dt=4e^{-2/3}$$

$$E(X^2)=18-34e^{-2/3}+4e^{-2/3}=18-30e^{-2/3}$$

$$Var(X)=E(X^2)-E^2(X)=\left(18-30e^{-2/3}\right)-\left(3-3e^{-2/3}\right)^2$$

We can easily verify that $E(X)=3-3e^{-2/3}$ is correct. Notice:

$$T+2=\min(T,2)+\max(T,2)$$

$$\Rightarrow\quad E(T+2)=E\left[\min(T,2)\right]+E\left[\max(T,2)\right]$$

We know that

$$E\left[\min(T,2)\right]=3-3e^{-2/3} \quad \text{(from this problem)}$$

$$E\left[\max(T,2)\right]=2+3e^{-2/3} \quad \text{(from the previous problem)}$$

$$E(T+2)=E(T)+2=3+2$$

So the equation $E(T+2)=E\left[\min(T,2)\right]+E\left[\max(T,2)\right]$ holds.

We can also check that $E(X^2)=18-30e^{-2/3}$ is correct.

$$T+2=\min(T,2)+\max(T,2)$$

$$\Rightarrow\qquad (T+2)^2=\left[\min(T,2)+\max(T,2)\right]^2$$

$$=\left[\min(T,2)\right]^2+\left[\max(T,2)\right]^2+2\min(T,2)\max(T,2)$$

$$\min(T,2)=\begin{cases} t & \text{if } t\le 2 \\ 2 & \text{if } t>2 \end{cases}, \qquad \max(T,2)=\begin{cases} 2 & \text{if } t\le 2 \\ t & \text{if } t>2 \end{cases}$$

$$\Rightarrow \qquad \min(T,2)\max(T,2)=2t$$

$$\Rightarrow \qquad (T+2)^2=\left[\min(T,2)\right]^2+\left[\max(T,2)\right]^2+2(2t)$$

$$\Rightarrow \qquad E(T+2)^2=E\left[\min(T,2)+\max(T,2)\right]^2$$

$$=E\left[\min(T,2)\right]^2+E\left[\max(T,2)\right]^2+E\left[2(2t)\right]$$

$$E(T+2)^2=E\left(T^2+4t+4\right)=E\left(T^2\right)+4E(t)+4=2\left(3^2\right)+4(3)+4=34$$

$$E\left[\min(T,2)\right]^2=18-30e^{-2/3} \quad \text{(from this problem)}$$

$$E\left[\max(T,2)\right]^2=4+30e^{-2/3} \quad \text{(from previous problem)}$$

$$E\left[2(2t)\right]=4E(t)=4(3)=12$$

$$E\left[\min(T,2)\right]^2+E\left[\max(T,2)\right]^2+E\left[2(2t)\right]$$
$$=18-30e^{-2/3}+4+30e^{-2/3}+12=34$$

So the equation $E(T+2)^2=E\left[\min(T,2)+\max(T,2)\right]^2$ holds.

Problem 8

An insurance company sells an auto insurance policy that covers losses incurred by a policyholder, subject to a deductible of \$100 and a maximum payment of \$300. Losses incurred by the policyholder are exponentially distributed with a mean of \$200. Find the expected payment made by the insurance company to the policyholder.

Solution

Let X =losses incurred by the policyholder. X is exponentially distributed with a mean of 200, $f(x)=\dfrac{1}{200}e^{-x/200}$.

Let Y =claim payment by the insurance company.

$$Y=\begin{cases} 0, & \text{if } X\le 100 \\ X-100, & \text{if } 100\le X\le 400 \\ 300, & \text{if } X\ge 400 \end{cases}$$

$$E(Y)=\int_0^{+\infty} y(x)\,f(x)\,dx=\int_0^{100} 0\,f(x)\,dx+\int_{100}^{400}(x-100)\,f(x)\,dx+\int_{400}^{+\infty}300\,f(x)\,dx$$

$$\int_0^{100} 0\,f(x)\,dx=0$$

$$\int_{100}^{400}(x-100)\,f(x)\,dx=\int_{100}^{400} xf(x)\,dx-\int_{100}^{400}100\,f(x)\,dx$$

$$\int_{100}^{400} xf(x)\,dx=(100+200)e^{-100/200}-(400+200)e^{-400/200}=300e^{-1/2}-600e^{-2}$$

$$\int_{100}^{400}100\,f(x)\,dx=100\left(e^{-100/200}-e^{-400/200}\right)=100\left(^{-1/2}-e^{-2}\right)$$

$$\int_{400}^{+\infty}300\,f(x)\,dx=300e^{-400/200}=300e^{-2}$$

Then we have

$$E(X)=300e^{-1/2}-600e^{-2}-100\left(e^{-1/2}-e^{-2}\right)+300e^{-2}=200\left(e^{-1/2}-e^{-2}\right)$$

Alternatively, we can use the shortcut developed in Chapter 20:

$$E(X)=\int_d^{d+L}\Pr(X>x)\,dx=\int_{100}^{100+300}e^{-x/200}dx=200\left[e^{-x/200}\right]_{400}^{100}=200\left(e^{-1/2}-e^{-2}\right)$$

Problem 9

An insurance policy has a deductible of 3. Losses are exponentially distributed with mean 10. Find the expected non-zero payment by the insurer.

Solution

Let X represent the losses and Y the payment by the insurer. Then $Y=0$ if $X\le 3$; $Y=X-3$ if $X>3$. We are asked to find $E\left(Y\middle|Y>0\right)$.

$$E\left(Y\middle|Y>0\right)=E\left(X-3\middle|X>3\right)$$

$X-3\,|\,X>3$ is an exponential random variable with the identical mean of 10. So $E\left(X-3\middle|X>3\right)=E(X)=10$.

Generally, if X is an exponential loss random variable with mean θ, then for any positive deductible d

$$E(X-d|X>d)=E(X)=\theta,\quad E(X|X>d)=E(X-d|X>d)+d=\theta+d$$

Problem 10

Claims are exponentially distributed with a mean of \$8,000. Any claim exceeding \$30,000 is classified as a big claim. Any claim exceeding \$60,000 is classified as a super claim.

Find the expected size of big claims and the expected size of super claims.

Solution

This problem tests your understanding that the exponential distribution lacks memory.
Let X represents claims. X is exponentially distributed with a mean of θ=8,000.
Let Y =big claims, Z =super claims.

$$\begin{aligned}E(Y)&=E(X|X>30,000)=E(X-30,000|X>30,000)+30,000\\&=E(X)+30,000=\theta+30,000=38,000\end{aligned}$$

$$\begin{aligned}E(Z)&=E(X|X>60,000)=E(X-60,000|X>60,000)+60,000\\&=E(X)+60,000=\theta+60,000=68,000\end{aligned}$$

Problem 11

Evaluate $\int_2^{+\infty}(x^2+x)e^{-x/5}dx$.

Solution

$$\int_2^{+\infty}(x^2+x)e^{-x/5}dx=5\int_2^{+\infty}(x^2+x)\left(\frac{1}{5}e^{-x/5}\right)dx=5\int_2^{+\infty}x^2\left(\frac{1}{5}e^{-x/5}\right)dx+5\int_2^{+\infty}x\left(\frac{1}{5}e^{-x/5}\right)dx$$

$$\int_2^{+\infty}x^2\left(\frac{1}{5}e^{-x/5}\right)dx=\left[5^2+(5+2)^2\right]e^{-2/5},\qquad\int_2^{+\infty}x\left(\frac{1}{5}e^{-x/5}\right)dx=(5+2)e^{-2/5}$$

$$\Rightarrow\qquad\int_2^{+\infty}(x^2+x)e^{-x/5}dx=5\left[5^2+((5+2)^2+(5+2)\right]e^{-2/5}=405e^{-2/5}$$

Problem 12 (two exponential distributions competing)

You have a car and a van. The time-to-failure of the car and the time-to-failure the van are two independent exponential random variables with mean of 8 years and 4 years respectively.

Calculate the probability that the car dies before the van.

Solution

Let X and Y represent the time-to-failure of the car and the time-to-failure of the van respectively. We are asked to find $P(X < Y)$.

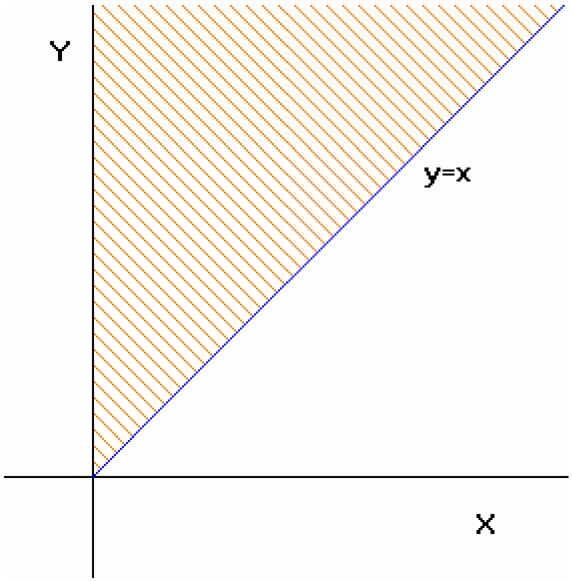
X and Y are independent exponential random variables with mean of 8 and 4 respectively. Their pdf is:

$$f_X(x) = \frac{1}{8}e^{-x/8}, \quad F_X(x) = 1 - e^{-x/8}, \text{ where } x \geq 0$$

$$f_Y(y) = \frac{1}{4}e^{-y/4}, \quad F_Y(y) = 1 - e^{-y/4}, \text{ where } y \geq 0$$

Method #1 X and Y have the following joint pdf:

$$f_{X,Y}(x, y) = f_X(x)\, f_Y(y) = \frac{1}{8}e^{-x/8}\left(\frac{1}{4}e^{-y/4}\right)$$



The shaded area is $x \geq 0$, $y \geq 0$, and $x < y$.

$$P(X < Y) = \iint_{shaded\ Area} f(x, y)\,dxdy = \int_0^{+\infty} \int_x^{+\infty} f(x, y)\,dydx = \int_0^{+\infty} \int_x^{+\infty} \frac{1}{8} e^{-x/8} \left(\frac{1}{4} e^{-y/4} \right) dydx$$

$$= \int_0^{+\infty} \frac{1}{8} e^{-x/8} \left(e^{-x/4} \right) dx = \int_0^{+\infty} \frac{1}{8} e^{-3x/8} dx = \frac{1}{3}$$

Method 2

$$P(X < Y) = \int_0^{+\infty} P(x < X \leq x + dx) P(Y > x + dx)$$

The above equation says that to that find $P(X < Y)$, we first fix X at a tiny interval $(x, x + dx]$. Next, we set $Y > x + dx$. This way, we are guaranteed that $X < Y$ when X falls in the interval $(x, x + dx]$. To find $P(X < Y)$ when X falls $[0, +\infty]$, we simply integrate $P(x < X \leq x + dx) P(Y > x + dx)$ over the interval $[0, +\infty]$.

$$P(x < X \leq x + dx) = f(x)\,dx = \frac{1}{8} e^{-x/8} dx$$

$$P(Y > x + dx) = P(Y > x) \qquad \text{because } dx \text{ is tiny}$$

$$= 1 - F_Y(x) = 1 - \left(1 - e^{-x/4}\right) = e^{-x/4}$$

$$P(X<Y)=\int_0^{+\infty} f_X(x)P(Y>x)=\int_0^{+\infty}\left(\frac{1}{8}e^{-x/8}dx\right)e^{-x/4}=\frac{1}{8}\int_0^{+\infty} e^{-\left(\frac{1}{8}+\frac{1}{4}\right)x}dx=\frac{\frac{1}{8}}{\frac{1}{8}+\frac{1}{4}}=\frac{1}{3}$$

This is the intuitive meaning behind the formula $\frac{\frac{1}{8}}{\frac{1}{8}+\frac{1}{4}}$. In this problem, we have a car and a van. The time-to-failure of the car and the time-to-failure the van are two independent exponential random variables with mean of 8 years and 4 years respectively. So on average a car failure arrives at the speed of 1/8 per year; van failure arrives at the speed of ¼ per year; and total failure (for cars and vans) arrives a speed of $\left(\frac{1}{8}+\frac{1}{4}\right)$ per year. Of the total failure, car failure accounts for $\frac{\frac{1}{8}}{\frac{1}{8}+\frac{1}{4}}=\frac{1}{3}$ of the total failure.

With this intuitive explanation, you should easily memorize the following shortcut:

In general, if X and Y are two independent exponential random variables with parameters of λ_1 and λ_2 respectively:

$$f_X(x)=\lambda_1 e^{-\lambda_1 x} \quad \text{and} \quad f_Y(y)=\lambda_2 e^{-\lambda_2 y}$$

Then $P(X<Y)=\int_0^{+\infty} f_X(x)P(Y>x)dx=\int_0^{+\infty}\lambda_1 e^{-\lambda_1 x}e^{-\lambda_2 x}dx=\lambda_1\int_0^{+\infty}e^{-(\lambda_1+\lambda_2)x}dx=\frac{\lambda_1}{\lambda_1+\lambda_2}$

Similarly, $P(Y>X)=\frac{\lambda_2}{\lambda_1+\lambda_2}$.

Now you see that $P(X<Y)+P(Y>X)=\frac{\lambda_1}{\lambda_1+\lambda_2}+\frac{\lambda_2}{\lambda_1+\lambda_2}=1$. This means that $P(X=Y)=0$. To see why $P(X=Y)=0$, please note that $X=Y$ is a line in the 2-D plane. A line doesn't have any area (i.e. the area is zero). If you integrate the joint pdf over a line, the result is zero.

If you have trouble understanding why $P(X=Y)=0$, you can think of probability in a 2-D plane as a volume. You can think of the joint pdf in a 2-D plane as the height function. In order to have a volume, you must integrate the height function over an area. A line doesn't have any area. Consequently, it doesn't have any volume.

Problem 13 (Sample P #90, also May 2000 #10)

An insurance company sells two types of auto insurance policies: Basic and Deluxe. The time until the next Basic Policy claim is an exponential random variable with mean two days. The time until the next Deluxe Policy claim is an independent exponential random variable with mean three days.

What is the probability that the next claim will be a Deluxe Policy claim?
(A) 0.172
(B) 0.223
(C) 0.400
(D) 0.487
(E) 0.500

Solution

Let T^B = time until the next Basic policy is sold. T^B is exponential random variable with $\lambda^B = \dfrac{1}{\theta^B} = \dfrac{1}{2}$.

Let T^D = time until the next Deluxe policy is sold. T^D is exponential random variable with $\lambda^D = \dfrac{1}{\theta^D} = \dfrac{1}{3}$.

"The next claim is a Deluxe policy" means that $T^D < T^B$.

$$P\left(T^D < T^B\right) = \frac{T^D}{T^D + T^B} = \frac{\frac{1}{3}}{\frac{1}{3}+\frac{1}{2}} = \frac{2}{5} = 0.4$$

Homework for you: #3 May 2000; #9, #14, #34 Nov 2000; #20 May 2001; #35 Nov 2001; #4 May 2003.

Chapter 19 Poisson distribution

Poisson distribution is often used to model the occurrence of a random event that happens unevenly (in some time periods it happens more often than other time periods). The occurrences of many natural events can be approximately modeled as a Poisson distribution such as:

- The number of claims that happen in a given time interval (a month, a year, etc.)
- The number of hits at a website in a given time interval
- The number of customers who arrive at a store in a given time interval
- The number of phone calls (or e-mails) you get in a day
- The number of shark attacks in one summer (Professor David Kelton of Penn State University used Poisson distribution to model the number of shark attacks in Florida in one summer.)

To enhance your understanding of Poisson distribution, I recommend that you read the article about using Poisson distribution to model the number of shark attacks:

http://www.pims.math.ca/pi/issue4/page12-14.pdf

A random variable X has a Poisson distribution with parameter $\lambda > 0$ if its probability mass function is:

$$f(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,2,\dots,\infty$$

λ is the occurrence rate per unit.

Poisson distribution is a special case of binomial distribution. For a binomial distribution with parameter n and p, if we let

$n \to +\infty$, $p \to 0$, but $np \to \lambda$

then the binomial distribution becomes Poisson distribution. You can find the proof in many textbooks.

To memorize the mean and variance of Poisson distribution, remember that binominal distribution with parameter n and p has a mean of np and a variance of $np(1-p)$. Let $p \to 0$ but $np \to \lambda$, then we see that a Poisson distribution has mean of λ and variance of λ.

Previous Course 1 Problems on Poisson distribution were straightforward. The major thing to watch out for is that you will need to convert the given occurrence rate of a random event into the occurrence rate of the time horizon in an exam problem.

Sample Problems and Solutions

Problem 1

Customers walk into a store at an average rate of 20 per hour. Find the probability that

(1) no customers have arrived at the store in 10 minutes.
(2) no more than 4 customers have arrived at the store in 30 minutes.

Solution

Customers walk into a store randomly and unevenly. Poisson distribution can be used to model the number of customers who have arrived at the store in a given interval.

To find the probability distribution of the number of customers who have walked into the store in 10 minutes, we need first to convert the arrival rate per hour into the arrival rate per 10 minutes.

Because there are 6 ten-minute intervals in an hour, the average # of arrivals per 10 minutes = 20 /6 = 10/3.

Let k =# of customer arrivals in 10 minutes.

$$f(k) = \frac{(10/3)^k}{k!} e^{-10/3}, \quad k=0,1,2,...$$

$$f(0) = \frac{(10/3)^0}{0!} e^{-10/3} = e^{-10/3} \approx 0.0357$$

Let n = # of customer arrivals in 30 minutes. The average # of arrivals in 30 minutes, λ, is 20/2=10.

$$f(n) = \frac{(10)^n}{n!} e^{-10}, \; n=0,1,2,...$$

$$P(n \le 4) = \sum_{n=0}^{4} \frac{(10)^n}{n!} e^{-10} = e^{-10}\left(1 + \frac{10}{1} + \frac{10^2}{2} + \frac{10^3}{3!} + \frac{10^4}{4!}\right) = 0.0293$$

Problem 2

The number of typos in a book has a Poisson distribution with an average of 5 typos per 100 pages. What is the probability that you cannot find a typo in 50 pages?

Solution

Let n = # of typos in 50 pages.

$$f(n)=\frac{\lambda^n}{n!}e^{-\lambda}, \ n=0,1,2,...$$

λ =average # of typos in 50 pages = 5/2=2.5

$$f(0)=\frac{2.5^0}{0!}e^{-2.5}=8.21\%$$

Problem 3

A beach resort buys a policy to insure against loss of revenues due to major storms in the summer. The policy pays a total of \$50,000 if there is only one major storm during the summer, a total of \$100,000 if there are two major storms, and a total payment of \$200,000 if there are more than two major storms.

The number of major storms in one summer is modeled by a Poisson distribution with mean of 0.5 per summer.

Find
(1) the expected premium for this policy during one summer.
(2) the standard of deviation of the cost of providing this insurance for one summer.

Solution

Let N =# of major storms in one summer. N has Poisson distribution with the mean λ=0.5.

$$f(n)=\frac{\lambda^n}{n!}e^{-\lambda}, \quad n=0,1,2,...$$

Let X = payment by the insurance company to the beach resort. The expected premium is the expected cost of the insurance:

$$E\left[X(n)\right]=\sum_{n=0}^{+\infty}X(n)f(n)$$

We will use BA II Plus to find the mean and variance.

n	Payment $X(n)$	$f(n)$	$10{,}000\,f(n)$
0	0	$e^{-0.5}$=0.6065	6,065
1	50,000	$0.5e^{-0.5}$=0.3033	3,033
2	100,000	$\frac{0.5^2}{2}e^{-0.5}$=0.0758	758
3+	200,000	1-(0.6065+0.3033+0.0758)=0.0144	144
Total			10,000

Using the BA II Plus 1-V Statistics Worksheet, you should get:

$$E(X)=25,625, \quad \sigma_X=37,889$$

Homework for you: #24, May 2000; #23 Nov 2000; #19 Nov 2001.

Chapter 20 Gamma distribution

A continuous random variable X has a gamma distribution if it has the following pdf:

$$f(x)=\frac{1}{\theta^{\alpha}\,\Gamma(\alpha)}x^{\alpha-1}e^{-x/\theta},\quad x\geq 0,\quad \alpha>0,\ \theta>0$$

Where α,θ are constants (called parameters).

$\Gamma(\alpha)$ is defined as

$$\Gamma(\alpha)=\int_0^{+\infty}t^{\alpha-1}e^{-t}\,dt\quad (\alpha>0)$$

$\Gamma(\alpha)$ is called the gamma **function** (pay attention here -- $\Gamma(\alpha)$ is not called gamma distribution). If α is a positive integer, then $\Gamma(\alpha)=(\alpha-1)!$

Gamma distribution can be very complex, especially when α is not an integer. For example, if $\alpha=1/2$, then the gamma pdf becomes:

$$f(x)=\frac{1}{\theta^{1/2}\,\Gamma(1/2)}x^{-1/2}e^{-x/\theta},\quad x\geq 0,\ \theta>0$$

Then finding just $\Gamma\left(\frac{1}{2}\right)$ can be very challenging:

$$\Gamma\left(\frac{1}{2}\right)=\int_0^{+\infty}t^{-\frac{1}{2}}e^{-t}\,dt=\int_0^{+\infty}x^{-1}e^{-x^2}\,dx^2=2\int_0^{+\infty}e^{-x^2}dx\quad \left(\text{let } t=x^2\right)$$

We must use polar transformation to find (details are not important and hence not shown here):

$$2\int_0^{+\infty}e^{-x^2}dx=\sqrt{\pi}\quad\Rightarrow\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$$

You might want to memorize (no need to learn how to prove these) the following basic facts about the gamma function:

$$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$$

Whether α is an integer or not, as long as $\alpha>0$, then

$$\Gamma(\alpha)=\int_0^{+\infty} t^{\alpha-1}e^{-t}\,dt \text{ exists,}$$

$$\Gamma(\alpha+1)=\alpha\,\Gamma(\alpha).$$

Though gamma distribution can be very complex, here is the good news: most likely you will be tested on a simplified version of gamma distribution where α is a positive integer.

Let's focus only on gamma distributions where α is a positive integer. We'll change α to n.

Simplified gamma distribution (most likely to be tested)

$$f(x)=\frac{1}{\theta^n(n-1)!}x^{n-1}e^{-x/\theta},\quad x\geq 0$$

To help us remember this complex pdf (so we can quickly and correctly write it out during the exam), we rewrite the pdf into:

$$f(x)=\lambda\times\underbrace{\frac{e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!}}_{\substack{\text{Possion probability of}\\ \text{having } (n\text{-}1) \text{ events}\\ \text{during interval } [0,x]}},\text{ where }\lambda=\frac{1}{\theta}$$

Or $$f(x)=\frac{1}{\theta}\underbrace{e^{-x/\theta}\frac{(x/\theta)^{n-1}}{(n-1)!}}_{\substack{\text{Poisson probability of}\\ \text{having } (n\text{-}1) \text{ events}\\ \text{during interval } [0,x]}}$$

In other words, we can express the gamma pdf as a multiple of Poisson pdf. The multiple is λ.

If $n=1$, then gamma distribution becomes exponential distribution.

One of the easiest ways to understand this simplified gamma distribution is to relate it to exponential distribution. While exponential distribution models the time elapsed until one random event occurs, **gamma distribution models the time elapsed until n random events occur**. You can find the proof of this in many textbooks.

Having n random events is the same as having a series of one random event. If a machine malfunctions three times during the next hour, the machine is really having three separate malfunctions within the next hour. As such, gamma distribution is really the sum of n identically distributed exponential random variables.

Let $T_1, T_2, ..., T_n$ be independent identically distributed exponential variables with parameter (mean) θ,

then $X = T_1 + T_2 + ... + T_n$ have gamma distribution with pdf

$$f(x) = \lambda \times \underbrace{e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!}}_{\substack{\text{Possion probability of} \\ \text{having } (n\text{-}1) \text{ events} \\ \text{during } [0,x]}}, \text{ where } \lambda = \frac{1}{\theta}$$

Once you understand this, you should have no trouble finding the mean and variance of X.

$$E(X) = E(T_1 + T_2 + ... + T_n) = E(T_1) + E(T_2) + ... + E(T_n) = n\theta$$
$$Var(X) = Var(T_1 + T_2 + ... + T_n) = Var(T_1) + Var(T_2) + ... + Var(T_n) = n\theta^2$$

We are not quite done yet. To solve gamma distribution problems, we need a quick way of finding the cdf. Brute force integration

$$F(x) = \int_0^x f(t)dt = \int_0^x \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} dt$$

can be painful.

You will want to memorize the following formula:

$$\begin{aligned} F(x) &= \Pr(X \le x) = \Pr(\text{it takes time } x \text{ or less to have } n \text{ random events}) \\ &= \Pr(\text{the \# of events that occurred during } [0,x] = n, n+1, n+2, ... + \infty) \\ &= 1\text{- } \Pr(\text{the \# of events that occurred during } [0,x] = 0, 1, 2, ... n-1) \\ &= 1 \text{ - } \underbrace{e^{-\lambda x}\left[1 + \frac{\lambda x}{1!} + \frac{(\lambda x)^2}{2!} + ... + \frac{(\lambda x)^{n-1}}{(n-1)!}\right]}_{\text{Poisson distribution}} \end{aligned}$$

Let's walk through the formula above. $\Pr(X \le x)$ really means that it takes no longer than time x before n random events occur. The only way to make this happen is to have

$n, n+1, n+2, ...$or $+\infty$ events occur during $\left[0, x\right]$. The number of events that can occur during $\left[0, x\right]$ is a Poisson distribution, which we already know how to calculate.

Please note that we use λx instead of λ as the parameter for Poisson distribution. λ is the occurrence rate per unit of time (or per unit of something). We have a total of x units. Consequently, the occurrence rate during the time interval $\left[0, x\right]$ is λx. I mentioned this point in the chapter on Poisson distribution. Forgetting this leads to a wrong result.

Similarly, we have

$$
\begin{aligned}
\Pr(X > x) &= \Pr(\text{it takes longer than time } x \text{ to have } n \text{ random events}) \\
&= \Pr(\text{the \# of events that occurred during } [0, x] = 0, 1, 2, ...\text{or } n-1) \\
&= \underbrace{e^{-\lambda x}\left[1 + \frac{\lambda x}{1!} + \frac{(\lambda x)^2}{2!} + ... + \frac{(\lambda x)^{n-1}}{(n-1)!}\right]}_{\text{Poisson distribution}}
\end{aligned}
$$

Now we are ready to tackle gamma distribution problems.

Sample Problems and Solutions

Problem 1

For a gamma distribution with parameter $n = 10, \lambda = 2$, quickly write out the expression for pdf and for $F(2)$.

Solution

$$
f(x) = \lambda \times \underbrace{e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}}_{\substack{\text{Poisson probability of} \\ \text{having } (n\text{-}1) \text{ events} \\ \text{during } \left[0, x\right]}}, \quad \text{where } \lambda = \frac{1}{\theta}
$$

$$
f(x) = \lambda \times e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!} = 2e^{-2x}\frac{(2x)^{10-1}}{(10-1)!} = 2e^{-2x}\frac{(2x)^9}{9!}
$$

$F(2)=\Pr(X\leq 2)=\Pr(\text{it takes time length of 2 or less to have 10 random events})$

$=\Pr(\text{the \# of events that occurred during } [0,2]=10,11,12,...+\infty)$

$=1-\Pr(\text{the \# of events that occurred during } [0,x]=0,1,2,...9)$

$$=1-\underbrace{e^{-\lambda x}\left[1+\frac{\lambda x}{1!}+\frac{(\lambda x)^2}{2!}+...+\frac{(\lambda x)^9}{9!}\right]}_{\text{Poisson distribution}}=1-e^{-4}\left[1+\frac{4}{1!}+\frac{(4)^2}{2!}+...+\frac{(4)^9}{9!}\right]$$

Problem 2

Given that a gamma distribution has the following cdf:

$$F(x)=1-\underbrace{e^{-\lambda x}\left[1+\frac{\lambda x}{1!}+\frac{(\lambda x)^2}{2!}+...+\frac{(\lambda x)^{n-1}}{(n-1)!}\right]}_{\text{Poisson distribution}}$$

Show that the pdf is indeed as follows:

$$f(x)=\lambda\, e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}$$

Solution

You won't be asked to prove this in the actual exam, but knowing the connection between gamma $F(x)$ and $f(x)$ enhances your understanding

$$f(x)=\frac{d}{dx}F(x)=\frac{d}{dx}\left\{1-e^{-\lambda x}\left[1+\frac{\lambda x}{1!}+\frac{(\lambda x)^2}{2!}+...+\frac{(\lambda x)^{n-1}}{(n-1)!}\right]\right\}$$

$$=-\frac{d}{dx}e^{-\lambda x}\left[1+\frac{\lambda x}{1!}+\frac{(\lambda x)^2}{2!}+...+\frac{(\lambda x)^{n-1}}{(n-1)!}\right]$$

$$=-\left[1+\frac{\lambda x}{1!}+\frac{(\lambda x)^2}{2!}+...+\frac{(\lambda x)^{n-1}}{(n-1)!}\right]\frac{d}{dx}e^{-\lambda x}-e^{-\lambda x}\frac{d}{dx}\left[1+\frac{\lambda x}{1!}+\frac{(\lambda x)^2}{2!}+...+\frac{(\lambda x)^{n-1}}{(n-1)!}\right]$$

$$=\lambda e^{-\lambda x}\left[1+\frac{\lambda x}{1!}+\frac{(\lambda x)^2}{2!}+...+\frac{(\lambda x)^{n-1}}{(n-1)!}\right]-e^{-\lambda x}\left[\lambda+\lambda\frac{\lambda x}{1!}+...+\lambda\frac{(\lambda x)^{n-2}}{(n-2)!}\right]$$

$$=\lambda e^{-\lambda x}\left[1+\frac{\lambda x}{1!}+\frac{(\lambda x)^2}{2!}+...+\frac{(\lambda x)^{n-1}}{(n-1)!}\right]-\lambda e^{-\lambda x}\left[1+\frac{\lambda x}{1!}+...+\frac{(\lambda x)^{n-2}}{(n-2)!}\right]$$

$$=\lambda e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}$$

Now we are convinced that our formula for $F(x)$ is correct.

Problem 3

You are solving three math problems. On average, you solve one problem every 2 minutes. What's the probability that you will solve all three problems in 5 minutes?

Solution

Method 1 – Forget about gamma distribution and use Poisson distribution

Let n =the number of problems you solve in 5 minutes. n has a Poisson distribution:

$$f(n) = e^{-\lambda x}\frac{(\lambda x)^n}{n!} \text{ where } \lambda = \frac{1}{2}, \quad x = 5$$

Pr(solve all 3 problems in 5 minutes)= $\Pr(n = 3, 4, 5, ...\infty) = 1 - \Pr(n = 0, 1, 2)$

$$= 1\text{-}e^{-\lambda x}\left[1+\frac{\lambda x}{1!}+\frac{(\lambda x)^2}{2!}\right] = 1\text{-}e^{-5/2}\left[1+\frac{5/2}{1!}+\frac{(5/2)^2}{2!}\right] = 45.62\%$$

Method 2 – Use gamma distribution and do integration

Let T_1 = the # of minutes it takes you to solve the 1st problem (the 1st problem can be any problem you choose to solve first)

Let T_2 = the # of minutes it takes you to solve the 2nd problem
Let T_3 = the # of minutes it takes you to solve the 3rd problem

T_1, T_2, T_3 are exponentially distributed with mean $\theta = 2$.

Let X =the total # of minutes it takes you to solve all three problems.

$$X = T_1 + T_2 + T_3$$

X has gamma distribution with $n = 3, \lambda = \frac{1}{\theta} = \frac{1}{2}$ and the following pdf:

$$f(x) = \lambda e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!} = \frac{1}{2}e^{-x/2}\frac{(x/2)^2}{2!}$$

$$\Pr(\text{solve all 3 problems in 5 minutes}) = \Pr(T \le 5)$$

$$= \int_0^5 f(t)dt = \int_0^5 \frac{1}{2} e^{-t/2} \frac{(t/2)^2}{2!} dt = 1 - \int_5^\infty \frac{1}{2} e^{-t/2} \frac{(t/2)^2}{2!} dt$$

$$= 1 - \int_5^\infty e^{-t/2} \frac{(t/2)^2}{2!} d\frac{t}{2} = 1 - \int_{5/2}^\infty e^{-x} \frac{x^2}{2!} dx = 1 - \frac{1}{2} \int_{5/2}^\infty x^2 e^{-x} dx \quad (\text{set } x = \frac{t}{2})$$

$$= 1 - \frac{1}{2} e^{-5/2} [1^2 + (1 + 5/2)^2]$$ (use the shortcut developed in the chapter on exponential distribution)

=45.62%

Method 3 – Use gamma distribution and avoid doing integration

We'll use the memorized formula of gamma cdf:

$$F(x) = 1 - \underbrace{e^{-\lambda x}\left[1 + \frac{\lambda x}{1!} + \frac{(\lambda x)^2}{2!} + ... + \frac{(\lambda x)^{n-1}}{(n-1)!}\right]}_{\text{Poisson distribution}}$$

We have $n = 3, \lambda = \frac{1}{\theta} = \frac{1}{2}$

$$F(x = 5) = 1 - e^{-5/2}\left[1 + \frac{5/2}{1!} + \frac{(5/2)^2}{2!}\right] = 45.62\%$$

You might want to familiarize yourself with all three methods above.

Problem 3

Solve the following integration:

$$\int_0^{+\infty} e^{-\delta t} t^n dt$$

where δ is a non-negative constant and n is a positive integer.

Solution

The above integration is frequently tested on SOA exams.

$$\int_0^{+\infty} e^{-\delta t} t^n dt = \frac{1}{\delta^{n+1}} \int_0^{+\infty} e^{-\delta t} (\delta t)^n d(\delta t) = \frac{1}{\delta^{n+1}} \int_0^{+\infty} e^{-x} x^n dx \quad (\text{let } x = \delta t)$$

$$\int_0^{+\infty} e^{-x} x^n dx = \Gamma(n+1) = n! \quad (\text{this is a gamma function})$$

$$\Rightarrow \int_0^{+\infty} e^{-\delta t} t^n dt = \frac{n!}{\delta^{n+1}}, \text{ or } \quad \int_0^{+\infty} e^{-\delta t} \frac{t^n}{n!} dt = \frac{1}{\delta^{n+1}}$$

for $\delta > 0$ and a non-negative integer *n*.

Generally, for a positive integer n,

$$\int_0^{+\infty} e^{-t} t^n dt = \Gamma(n+1) = n!, \text{ where } \Gamma(n+1) \text{ is a gamma function.}$$

Alternatively,

$$\int_0^{+\infty} e^{-\delta t} t^n dt = \frac{[(n+1)-1]!}{\delta^{n+1}} \int_0^{+\infty} \underbrace{\delta e^{-\delta t} \frac{(\delta t)^{(n+1)-1}}{[(n+1)-1]!}}_{\substack{\text{gamma pdf with} \\ \text{parameter } \delta \text{ and } n+1}} dt = \frac{[(n+1)-1]!}{\delta^{n+1}} = \frac{n!}{\delta^{n+1}}$$

(The integration of a gamma pdf over $[0, +\infty]$ is one). This gives us:

$$\boxed{\int_0^{+\infty} e^{-\delta t} t^n dt = \frac{n!}{\delta^{n+1}},}$$

To help memorize the above equation, let's rewrite the above formulas as

$$\boxed{\int_0^{+\infty} e^{-\delta t} \frac{(\delta t)^n}{n!} d(\delta t) = 1 \text{ for } \delta > 0 \text{ and a non-negative integer } n.}$$

Notice that $e^{-\delta t} \frac{(\delta t)^n}{n!}$ in the integration sign is the Poisson probability mass function $P(N = n)$ with parameter $\lambda = \delta t$. Setting $\lambda = \delta t$, we have:

$$\int_0^{+\infty} \underbrace{e^{-\lambda}\frac{\lambda^n}{n!}}_{\text{Poisson probability mass function}} d\lambda = 1.$$

So you just need to memorize:

$$\boxed{\int_0^{+\infty} \underbrace{e^{-\lambda}\frac{\lambda^n}{n!}}_{\text{Poisson probability mass function}} d\lambda = 1 \quad \text{or} \quad \int_0^{+\infty} \underbrace{e^{-\delta t}\frac{(\delta t)^n}{n!}}_{\text{Poisson probability mass function}} d(\delta t) = 1}$$

If we set $\lambda = x$ and $n = 1$, we get $\int_0^{+\infty} e^{-x} x dx = 1$. $\int_0^{+\infty} e^{-x} x dx$ is the mean of an exponential random variable with parameter $\theta = 1$. So $\int_0^{+\infty} e^{-x} x dx = 1$ is correct.

If we set $\lambda = x$ and $n = 2$, we get $\int_0^{+\infty} e^{-x}\frac{x^2}{2!} dx = 1$ or $\int_0^{+\infty} e^{-x} x^2 dx = 2$. If we do integration-by-parts, we get:

$$\int_0^{+\infty} e^{-x} x^2 dx = -\int_0^{+\infty} x^2 de^{-x} = -\left[x^2 e^{-x}\right]_0^{+\infty} + \int_0^{+\infty} e^{-x} dx^2 = \int_0^{+\infty} e^{-x} dx^2 = 2\int_0^{+\infty} e^{-x} x dx = 2$$

So we know that $\int_0^{+\infty} e^{-x}\frac{x^2}{2!} dx = 1$ is correct.

In the future, if you see $\int_0^{+\infty} e^{-x} x^n dx$, immediately change $e^{-x} x^n$ to a Poisson probability mass function: $\int_0^{+\infty} e^{-x} x^n dx = n!\int_0^{+\infty} e^{-x}\frac{x^n}{n!} dx = n!$

Problem 4

Solve the following integration:

$$\int_0^{+\infty} e^{-3t} t^5 dt$$

Solution

Let's change $e^{-3t}t^5$ into a Poisson probability mass function:

$$\int_0^{+\infty} e^{-3t}t^5 dt = \frac{5!}{3^6}\int_0^{+\infty} e^{-(3t)}\frac{(3t)^5}{5!}d(3t) = \frac{5!}{3^6}$$

Problem 5

Given that $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$, find $\Gamma\left(\frac{11}{2}\right)$

Solution

Generally, $\Gamma(\alpha+1)=\alpha\,\Gamma(\alpha)$ where $\alpha>0$.

$$\Gamma\left(\frac{11}{2}\right)=\Gamma\left(\frac{9}{2}+1\right)=\frac{9}{2}\Gamma\left(\frac{9}{2}\right)$$

$$\Gamma\left(\frac{9}{2}\right)=\Gamma\left(\frac{7}{2}+1\right)=\frac{7}{2}\Gamma\left(\frac{7}{2}\right)$$

$$\Gamma\left(\frac{7}{2}\right)=\Gamma\left(\frac{5}{2}+1\right)=\frac{5}{2}\Gamma\left(\frac{5}{2}\right)$$

$$\Gamma\left(\frac{5}{2}\right)=\Gamma\left(\frac{3}{2}+1\right)=\frac{3}{2}\Gamma\left(\frac{3}{2}\right)$$

$$\Gamma\left(\frac{3}{2}\right)=\Gamma\left(\frac{1}{2}+1\right)=\frac{1}{2}\Gamma\left(\frac{1}{2}\right)$$

$$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$$

$$\Rightarrow\quad \Gamma\left(\frac{11}{2}\right)=\left(\frac{9}{2}\right)\left(\frac{7}{2}\right)\left(\frac{5}{2}\right)\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\left(\sqrt{\pi}\right)$$

Homework for you: Rework all the problems in this chapter.

Chapter 21 Beta distribution

Let's pick up where we left off on beta distribution in Chapter 7.

You are tossing a coin. Not knowing p, the success rate of heads showing up in one toss of the coin, you subjectively assume that p is uniformly distributed over $\left[0,1\right]$. Next, you do an experiment by tossing the coin $m+n$ times (where m,n are non-negative integers). You find that, in this experiment, m out of these $m+n$ tosses have heads.

Then the posterior probability of p is:

$$f\left(p\right)=\frac{p^m\left(1-p\right)^n}{\int_0^1 p^m\left(1-p\right)^n dp} \qquad \text{where } 0\le p\le 1;\ m,n \text{ are non-negative integers.}$$

The above distribution $f\left(p\right)$ is called beta distribution.

If we set $m=\alpha-1$ and $n=\beta-1$ where $\alpha>0$ and $\beta>0$, we have

$$f\left(p\right)=\frac{p^{\alpha-1}\left(1-p\right)^{\beta-1}}{\int_0^1 p^{\alpha-1}\left(1-p\right)^{\beta-1} dp}, \quad \text{where } 0\le p\le 1$$

This is a generalized beta distribution.

Key points to remember:

General beta distribution:

pdf:

$$f(p)=\frac{p^{\alpha-1}(1-p)^{\beta-1}}{\int_0^1 p^{\alpha-1}(1-p)^{\beta-1}dp}=\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}$$

where $\alpha>0$, $\beta>0$, $0\le p\le 1$

$$B(\alpha,\beta)=\int_0^1 p^{\alpha-1}(1-p)^{\beta-1}dp=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad \text{(Beta function)}$$

Cdf:

$$F(p)=\int_0^p f(x)dx=\frac{1}{B(\alpha,\beta)}\int_0^p x^{\alpha-1}(1-x)^{\beta-1}dx=\frac{B_x(\alpha,\beta)}{B(\alpha,\beta)}=I_x(\alpha,\beta)$$

Where

$B_x(\alpha,\beta)=\int_0^p x^{\alpha-1}(1-x)^{\beta-1}dx$ is called the incomplete beta function

$I_x(\alpha,\beta)$ is called the incomplete beta function ratio.

Mean and variance – good if you can memorize them:

$$E(P)=\frac{\alpha}{\alpha+\beta},\ Var(P)=\left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\beta}{\alpha+\beta}\right)\left(\frac{1}{\alpha+\beta+1}\right)$$

Simplified Beta distribution (likely to be on the exam): where α and β are positive integers. Let $m=\alpha-1$ and $n=\beta-1$.

pdf:

$$f(p)=\frac{p^m(1-p)^n}{\int_0^1 p^m(1-p)^n dp}=\underbrace{(m+n+1)}_{\text{\# of trials + 1}}\times\underbrace{C_{m+n}^m\, p^m(1-p)^n}_{\substack{\text{binomial distribution}\\ m\text{ successes}\\ n\text{ failures}}}$$

where $0\le p\le 1$ and m,n are non-negative integers

cdf: Two methods

Method 1 Do the following integration:

$$F(p)=\int_0^p f(x)dx=\int_0^p (m+n+1)C_{m+n}^m\, x^m(1-x)^n dx$$

Method 2 $F(p)$=Binomial probability of having more than m successes in $m+n+1$ trials, where the success rate is p in one trial.

Or $$F(p)=\sum_{k=m+1}^{m+n+1} C_{m+n+1}^k\, p^k(1-p)^{m+n+1-k}$$

The proof of Method 2 is complex. And there's no intuitive explanation for it. Just memorize it.

Integration shortcut -- comes in handy in the heat of the exam:

$$\int_0^1 p^m(1-p)^n dp=\frac{1}{(m+n+1)\,C_{m+n}^m}$$

This is how we get

$$f(p)=(m+n+1)C_{m+n}^{m}\ p^{m}(1-p)^{n} \text{ and } \int_0^1 p^{m}(1-p)^{n}dp=\frac{1}{(m+n+1)\,C_{m+n}^{m}}.$$

Proof.

$$f(p)=\frac{p^{m}(1-p)^{n}}{\int_0^1 p^{m}(1-p)^{n}dp}=\frac{p^{m}(1-p)^{n}}{B(m+1,n+1)},$$

$$B(m+1,n+1)=\frac{\Gamma(m+1)\Gamma(n+1)}{\Gamma(m+n+2)}=\frac{m!n!}{(m+n+1)!}$$

$$\Rightarrow\quad f(p)=\frac{(m+n+1)!}{m!n!}p^{m}(1-p)^{n}=(m+n+1)C_{m+n}^{m}\ p^{m}(1-p)^{n}$$

$$\int_0^1 f(p)dp=1\quad\Rightarrow\int_0^1(m+n+1)C_{m+n}^{m}\ p^{m}(1-p)^{n}dp=1$$

$$\Rightarrow\int_0^1 p^{m}(1-p)^{n}dp=\frac{1}{(m+n+1)\,C_{m+n}^{m}}$$

Sample problem and solutions

Problem 1

A random variable X (where $0\le X\le 1$) has the following pdf:

$f(x)=k\,x^{5}(1-x)^{2}$, where k is a constant.

Find $k, E(X), Var(X)$.

Solution

Method 1: Using memorized formulas for general beta distribution

X has beta distribution with parameters $\alpha = 6, \beta = 3$.

$$k = \frac{1}{B(\alpha, \beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\Gamma(6+3)}{\Gamma(6)\Gamma(3)} = \frac{\Gamma(9)}{\Gamma(6)\Gamma(3)} = \frac{8!}{5!2!} = 168$$

$$E(P) = \frac{\alpha}{\alpha+\beta} = \frac{6}{6+3} = \frac{2}{3}$$

$$Var(P) = \left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\beta}{\alpha+\beta}\right)\left(\frac{1}{\alpha+\beta+1}\right) = \left(\frac{6}{6+3}\right)\left(\frac{3}{6+3}\right)\left(\frac{1}{6+3+1}\right) = \frac{1}{45}$$

Method 2 – do integration

First, we'll find k.

$$\int_0^1 f(x)dx = 1 \quad \Rightarrow \int_0^1 k\, x^5 (1-x)^2 dx = 1 \quad \Rightarrow k \int_0^1 x^5 (1-x)^2 dx = 1$$

$$\int_0^1 x^5 (1-x)^2 dx = \frac{1}{(5+2+1)\, C_{5+2}^5} = \frac{1}{8\, C_7^5} = \frac{1}{8\, C_7^2} \quad \Rightarrow k = 8\, C_7^2$$

Alternatively, we have # of success=5, # of failures=2.

$$k = (\#\ \text{of trials} + 1)\, C_{\#\text{ of trials}}^{\#\text{ of successes}} = (5+2+1)\, C_{5+2}^{5} = 8C_7^5 = 8C_7^2 = 168$$

$$E(X) = \int_0^1 x f(x)dx = \int_0^1 k\, x^6 (1-x)^2 dx = k \int_0^1 x^6 (1-x)^2 dx$$

$$\int_0^1 x^6 (1-x)^2 dx = \frac{1}{(6+2+1)\, C_8^6} = \frac{1}{9\, C_8^2}$$

$$E(X)=k\int_0^1 x^6\,(1-x)^2\,dx=\frac{k}{9\,C_8^2}=\frac{8\,C_7^2}{9\,C_8^2}=\frac{\frac{8\,(7)(6)}{2!}}{\frac{9(8)(7)}{2!}}=\frac{6}{9}=\frac{2}{3}$$

$$E(X^2)=\int_0^1 x^2 f(x)dx=\int_0^1 k\,x^7\,(1-x)^2\,dx=k\int_0^1 x^7\,(1-x)^2\,dx$$

$$\int_0^1 x^7\,(1-x)^2\,dx=\frac{1}{(7+2+1)\,C_{7+2}^7}=\frac{1}{10\,C_9^2}$$

$$E(X^2)=k\int_0^1 x^7\,(1-x)^2\,dx=\frac{k}{10\,C_9^2}=\frac{8\,C_7^2}{10\,C_9^2}=\frac{\frac{8(7)(6)}{2!}}{\frac{10(9)(8)}{2!}}=\frac{7}{15}$$

$$Var(X)=E(X^2)-E^2(X)=\frac{7}{15}-\left(\frac{2}{3}\right)^2=\frac{1}{45}$$

Problem 2

The percentage of defective products in a batch of products, p, is assumed to be uniformly distributed in [0,1]. An engineer randomly chooses 50 items with replacement from this batch and discovers no defective products. Determine the posterior probability that more than 5% of the products in this batch are defective.

Solution

If 50 items are sampled with replacement, then the number of defective items found in this sample is a binomial distribution (if it is a sample without replacement, you'll get a hypergeometric distribution). Given our interpretation of Beta distribution, the posterior probability of p has Beta distribution:

$$f\left(p\right)=\frac{p^m\left(1-p\right)^n}{\int_0^1 p^m\left(1-p\right)^n dp}=\underbrace{(m+n+1)}_{\#\text{ of trials + 1}}\times\underbrace{C_{m+n}^m\;p^m(1-p)^n}_{\substack{\text{binomial distribution}\\ m\text{ successes}\\ n\text{ failures}}}$$

In this problem, $m=0, n=50$.

$$f\left(p\right)=(m+n+1)C_{m+n}^m\;p^m(1-p)^n=(0+50+1)C_{0+50}^0\;p^0(1-p)^{50}=51(1-p)^{50}$$

$$\Pr(p > 5\%) = \int_{5\%}^{1} f(p)\,dp = \int_{5\%}^{1} 51(1-p)^{50}\,dp = -\int_{5\%}^{1} d(1-p)^{51} = -\left[(1-p)^{51}\right]_{5\%}^{1}$$
$$= (.95)^{51} = 7.3\%$$

Alternatively,

$$\Pr(p > 5\%) = 1 - \Pr(p \le 5\%) = 1 - F(5\%)$$

$F(5\%)$=Binomial probability of having more than zero success in 51 trials, where the success rate is 5% per trial.

$\Rightarrow \quad 1 - F(5\%)$=having zero success in 51 trials=$(.95)^{51} = 7.3\%$.

Problem 3

You are tossing a coin. Not knowing p, the success rate of heads showing up in one toss of the coin, you subjectively assume that p has the following distribution over $[0,1]$:

$$f(p) = \frac{1}{3}p^2$$

Next, you do an experiment by tossing the coin 10 times. You find that, in this experiment, 2 out of these 10 tosses have heads.

Find the mean of the posterior probability of p.

Solution

Using Bayes' Theorem, we know the posterior probability is:

$$\underbrace{\underbrace{k}_{\text{scaling factor}} \underbrace{\frac{1}{3}p^2}_{\substack{\text{before-event} \\ \text{group size}}} \underbrace{C_{10}^{2}\, p^2 (1-p)^8}_{\substack{\text{the group's probability} \\ \text{to have 2 heads out of 10 tosses}}}}_{\text{After-event size of the groups}}$$

So the posterior probability has the following form:

constant $\times p^4 (1-p)^8$, where $0 \le p \le 1$.

Without knowing exactly what the constant is, we see that the posterior probability is a Beta distribution with $\alpha = 5$ and $\beta = 9$.

Next, we can simply use the following memorized formula:

$$E(P) = \frac{\alpha}{\alpha + \beta} = \frac{5}{5+9} = \frac{5}{14}$$

Alternatively,

$f(p) = c\, p^4 (1-p)^8$, where c is a constant.

$$\int_0^1 f(p)dp = 1 \quad \Rightarrow \int_0^1 c\, p^4 (1-p)^8 dp = 1, \quad c = \frac{1}{\int_0^1 p^4 (1-p)^8 dp} = (4+8+1)C_{4+8}^4 = 13C_{12}^4$$

$$E(P) = \int_0^1 pf(p)\,dp = \int_0^1 p\, 13\, C_{12}^4 p^4 (1-p)^8\, dp = 13\, C_{12}^4 \int_0^1 p^5 (1-p)^8\, dp$$

$$= \frac{13\, C_{12}^4}{(5+8+1)C_{5+8}^5} = \left(\frac{13}{14}\right)\left(\frac{\frac{12(11)(10)(9)}{4!}}{\frac{13(12)(11)(10)(9)}{5!}}\right) = \frac{5}{14}$$

Problem 4

A random variable X (where $0 \le X \le 1$) has the following pdf:

$f(x) = k\, x^3 (1-x)^2$, where k is a constant.

Find $\Pr\left(X \le \frac{1}{3}\right) = F\left(\frac{1}{3}\right)$.

Solution

X has a simplified beta distribution with parameters $m=3$ and $n=2$.

Method 1 – do integration.

$$\Pr\left(X \le \frac{1}{3}\right) = F\left(\frac{1}{3}\right) = \int_0^{\frac{1}{3}} f(x)dx = \int_0^{\frac{1}{3}} k\, x^3 (1-x)^2 dx$$

$$k = (m+n+1)C_{m+n}^{m} = (3+2+1)C_{3+2}^{3} = 6C_5^3 = 6C_5^2 = (6)\frac{5(4)}{2!} = 60$$

$$F\left(\frac{1}{3}\right) = \int_0^{\frac{1}{3}} 60\, x^3 (1-x)^2 dx = \int_0^{\frac{1}{3}} 60\, x^3 (x^2-2x+1)dx = 60\int_0^{\frac{1}{3}} (x^5-2x^4+x^3)dx$$

$$= 60\int_0^{\frac{1}{3}} (x^5-2x^4+x^3)dx = 60\left[\frac{1}{6}x^6 - \frac{2}{5}x^5 + \frac{1}{4}x^4\right]_0^{\frac{1}{3}} \approx 10.0137\%$$

Method 2

$F\left(\frac{1}{3}\right)$ =Binomial probability of having more than $m=3$ successes in $m+n+1=3+2+1=6$ trials, where the success rate is $p=\frac{1}{3}$ in one trial.

$$F\left(\frac{1}{3}\right) = C_6^4\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2 + C_6^5\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right) + C_6^6\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^0$$

$$\approx 8.23045\% + 1.64609\% + 0.13717\% \approx 10.0137\%$$

Problem 5

A random variable X (where $0 \le X \le 1$) has the following pdf:

$f(x) = k\, x^2 (1-x)^{\frac{3}{2}}$, where k is a constant.

Find $\Pr\left(X \le \frac{1}{3}\right) = F\left(\frac{1}{3}\right)$.

Solution

X is a beta random variable with parameters $\alpha = 2+1=3$ and $\beta = \frac{3}{2}+1=\frac{5}{2}$

Because one parameter is a fraction, we cannot use binomial distribution to calculate $F\left(\frac{1}{3}\right)$ as we did in the previous problem. We have to do integration. In addition, we do not like $(1-x)^{\frac{3}{2}}$. Let's set $y=1-x$.

$$\Pr\left(X \le \frac{1}{3}\right) = \int_0^{\frac{1}{3}} k\, x^2 (1-x)^{\frac{3}{2}} dx = -\int_1^{\frac{2}{3}} k\,(1-y)^2\, y^{\frac{3}{2}} dy = \int_{\frac{2}{3}}^{1} k\, y^{\frac{3}{2}} (1-y)^2\, dy$$

$$k = \frac{\Gamma\left(3+\frac{5}{2}\right)}{\Gamma(3)\Gamma\left(\frac{5}{2}\right)} = \frac{\Gamma(5.5)}{\Gamma(3)\Gamma(2.5)} = \frac{(4.5)(3.5)(2.5)\Gamma(2.5)}{(2!)\Gamma(2.5)} = \frac{(4.5)(3.5)(2.5)}{2} = 19.6875$$

$$\int_{\frac{2}{3}}^{1} y^{\frac{3}{2}} (1-y)^2\, dy = \int_{\frac{2}{3}}^{1} y^{1.5}\left(y^2-2y+1\right) dy = \int_{\frac{2}{3}}^{1} \left(y^{3.5}-2y^{2.5}+y^{1.5}\right) dy$$

$$= \left[\frac{1}{4.5} y^{4.5} - \frac{2}{3.5} y^{3.5} + \frac{1}{2.5} y^{2.5}\right]_{\frac{2}{3}}^{1} \approx 0.804\%$$

$$\Pr\left(X \le \frac{1}{3}\right) \approx 19.6875(0.804\%) \approx 15.83\%$$

Final comment. If you are given the following beta distribution (both parameters are fractions)

$$f(x) = k\, x^{\frac{5}{2}} (1-x)^{\frac{3}{2}}$$

and you are asked to find $\Pr\left(X \le \frac{1}{3}\right) = F\left(\frac{1}{3}\right)$, then the integration is tough. To do the integration, you might want to try setting

$$x = \sin^2 t,\ y = \cos^2 t$$

This is too much work. It's unlikely that SOA will include this type of heavy calculus in the exam.

Homework for you: Rework all the problems in this chapter

Chapter 22 Weibull distribution

We know that a component's time to failure can be modeled using exponential distribution. Now let's consider a machine that has *n* components. The machine works as long as all the components are working; it stops working if at least one component stops working.
What is the probability distribution of the machine's time to failure?

Let's make two simplifying assumptions: (1) each component is independent of any other component, and (2) each component's time to failure is exponentially distributed with identical mean θ.

Let T_1 represent the first component's time to failure. T_1 is exponentially distributed with mean θ. $\Pr(T_1 > t) = e^{-t/\theta}$

Let T_2 represent the second component's time to failure. T_2 is exponentially distributed with mean θ. $\Pr(T_2 > t) = e^{-t/\theta}$

……

Let T_n represent the *n*-th component's time to failure. T_n is exponentially distributed with mean θ. $\Pr(T_n > t) = e^{-t/\theta}$

Let T represent the machine's time to failure. Then $T = \min(T_1, T_2, ..., T_n)$.

To derive the pdf of T, notice

$$\Pr(T > t) = \Pr\left[\min(T_1, T_2, ..., T_n) > t\right] = \Pr\left[(T_1 > t)\cap(T_2 > t)\cap...(T_n > t)\right]$$

Because $T_1, T_2, ..., T_n$ are independent, we have

$$\Pr\left[(T_1 > t)\cap(T_2 > t)\cap...(T_n > t)\right] = \Pr(T_1 > t)\Pr(T_2 > t)...\Pr(T_n > t) = \left(e^{-t/\theta}\right)^n$$

$$F(T) = \Pr(T \le t) = 1 - \Pr(T > t) = 1 - \left(e^{-t/\theta}\right)^n$$

$F(T)=1-\left(e^{-t/\theta}\right)^n$ is very close to Weibull distribution. However, there's one more step. Statisticians found out that if they change $F(T)=1-\left(e^{-t/\theta}\right)^n=1-e^{-nt/\theta}$ into $F(T)=1-e^{-(x/\theta)^\beta}$, the resulting distribution $F(T)=1-e^{-(x/\theta)^\beta}$ is much more useful. There's no good theoretical justification why $F(T)=1-\left(e^{-t/\theta}\right)^n=1-e^{-nt/\theta}$ needs to be changed to $F(T)=1-e^{-(x/\theta)^\beta}$. The key point is that people who use Weibull distribution don't care much where the cdf $F(T)=1-e^{-(x/\theta)^\beta}$ comes from. All they care is that this cdf is very flexible and they can easily fit their data into this cdf. This is pretty much all you need to know about the theories behind Weibull distribution.

If an object's time to failure, X, follows Weibull distribution, then its probability to fail by time x is:

$$\Pr(X \le x)=1-e^{-(x/\theta)^\beta} \quad \text{where } \theta>0, \beta>0.$$

Stated differently, if an object's time to failure, X, follows Weibull distribution, then its probability to survive time x is:

$$\Pr(X > x)=e^{-(x/\theta)^\beta} \quad \text{where } \theta>0, \beta>0.$$

θ is called the **scale parameter** and β is called the **shape parameter**.

Weibull distribution is widely used to describe the failure time of a machine (a car, a vacuum cleaner, light bulbs, etc.) which consists of several components and which fails to work it at least one component stops working.

Please note that many textbooks use the following notation:

$$\Pr(X > x)=e^{-(x/\alpha)^\beta}$$

I like to use $\Pr(X > x)=e^{-(x/\theta)^\beta}$ to help me remember that Weibull is just a complex version of exponential distribution.

What is so special about Weibull distribution? Weibull cdf can take a variety of shapes such as a bell curve, a U shape, a J shape, a roughly straight line, or some other shape. To get a feel for the Weibull pdf shape, go to

http://www.engr.mun.ca/~ggeorge/3423/demos/

and download the Weibull Probability Distribution Excel spreadsheet. This spreadsheet lets you enter parameters α and β. Then it displays the corresponding graphs for Weibull pdf and cdf.

You can play around with the spreadsheet. Enter different parameters in the spreadsheet and watch how Weibull pdf and cdf shapes change.

Because Weibull pdf can take a variety of shapes, Weibull distribution can fit a variety of data. This is why Weibull distribution has many applications. When you use Weibull distribution to fit data, you recognize the fact that your data may not fit into a neat bell curve, a neat exponential distribution, or other standard shape. By using Weibull distribution, you give your data a chance to "speak for themselves." In contrast, if you use a distribution with a fixed shape, say a normal distribution, you implicitly force your data to fit into a bell curve.

To solve Weibull distribution-related problems, remember the following key points:

1. Do not be scared. Weibull distribution sounds hard, but the calculation is simple.

2. The bare-bones formula you want to memorize about a Weibull random variable X (where $X \geq 0$) is $\Pr(X > x) = e^{-(x/\theta)^{\beta}}$, where $\theta > 0, \beta > 0$. Compare this with the formula for an exponential variable X: $\Pr(X > x) = e^{-x/\theta}$, and you see that **exponential distribution is just a simplified version of Weibull distribution by setting** $\beta = 1$**.**

3. About the pdf, mean and variance formula, you have two options. One option is to memorize the formulas; the other is to derive the mean and variance from the bare bones formula of $\Pr(X > x)$.

I recommend that you master both options. You should use Option 1 for the exam and use Option 2 as a backup (in case you forget the formulas).

Method 1 – memorize the formulas for $E(X), Var(X)$

Under Method 1, you still do not need to memorize the pdf. You can find the pdf using the formula:

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}\Pr(X \leq x) = \frac{d}{dx}\left[1 - \Pr(X > x)\right] = -\frac{d}{dx}\Pr(X > x)$$

$$\Rightarrow \quad f(x) = \frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1} e^{-(x/\theta)^{\beta}}$$

However, to find the mean and variance, you will want to memorize:

$$E(X^n)=\theta^n\,\Gamma\left(1+\frac{n}{\beta}\right),\text{ where } n=1,2,\ldots$$

If $\dfrac{n}{\beta}$ is an integer (likely to appear on the exam), then the above formula becomes

$$E(X^n)=\theta^n\left(\frac{n}{\beta}!\right)$$

Method 2 - derive $E(X),Var(X)$ **using general probability formulas.** Method 2 is not hard (I will show you how).

Derive pdf: (same as Method 1)

$$f(x)=\frac{d}{dx}F(x)=\frac{d}{dx}\Pr(X\le x)=\frac{d}{dx}\left[1-\Pr(X>x)\right]=-\frac{d}{dx}\Pr(X>x)$$

Derive mean and variance:

$$E(X^n)=\int_0^{+\infty}f(x)\,x^n dx$$

To simplify our integration, we use the following shortcut:

For a non-negative random variable X where $0\le X\le+\infty$,

$$E(X^n)=\int_0^{+\infty}\Pr(X>x)dx^n$$

Please note that this shortcut works for any non-negative variable with a lower bound of zero and an upper bound of positive infinity.

Proof.

$$f(x)=-\frac{d}{dx}\Pr(X>x)\quad\Rightarrow f(x)dx=-d\left[\Pr(X>x)\right]$$

$$E(X^n)=\int_0^{+\infty}f(x)x^n dx=\int_0^{+\infty}x^n f(x)dx=-\int_0^{+\infty}x^n d\left[\Pr(X>x)\right]$$

$$=-x^n\left[\Pr(X>x)\right]\Big|_0^{+\infty}+\int_0^{+\infty}\Pr(X>x)dx^n\quad\text{(integration by parts)}$$

$$-x^n\left[\Pr(X>x)\right]\Big|_0^{+\infty}=0^n\Pr(X>0)-\left(+\infty^n\right)\Pr\left(X>+\infty\right)=-\left(+\infty^n\right)\Pr\left(X>+\infty\right)$$

$\left(+\infty^{n}\right)\Pr\left(X>+\infty\right)$ cannot be ∞. If it is ∞, $E(X^{n})$ will be undefined. But if you are asked to calculate $E(X^{n})$, it must exist in the first place. So $\left(+\infty^{n}\right)\Pr\left(X>+\infty\right)$ must be zero.

If you feel uneasy about setting $\left(+\infty^{n}\right)\Pr\left(X>+\infty\right)$ to zero, use a human life as the random variable. Let X represent the number of years until the death of a human being. If you choose a large number such as $x=200$, then you will surely have $\Pr(X>200)=0$ because no one can live for 200 years. Clearly, we can set $x^{n}\Pr(X>x)\Big|_{+\infty}=0$. So we have $E\left(X^{n}\right)=\int_{0}^{+\infty}\Pr\left(X>x\right)dx^{n}$

Sample problems and solutions

Problem 1

For a Weibull random variable X with parameters $\theta=10, \beta=\frac{1}{4}$, find the pdf, $E(X), Var(X)$, 80th percentile, and Pr($X>20$).

Solution

We will first solve the problem with Method 2 (the hard way). If you can master this method you will do fine in the exam, even if you forget the formulas.

Method 2 – this is the "Let's focus on the basic formulas and derive the rest" approach.

For Weibull distribution, we always start from $\Pr\left(X>x\right)$ (which is called the survival function), not from $f(x)$. $\Pr\left(X>x\right)$ is lot easer to memorize than cdf.

Let's start with $\Pr\left(X>x\right)$. From here we can find everything else.

$$\Pr\left(X>x\right)=e^{-(x/\theta)^{\beta}}=e^{-\left(x/\theta\right)^{1/4}} \quad \text{where } x\geq 0$$

$$f(x)=-\frac{d}{dx}\Pr(X>x)=-\frac{d}{dx}e^{-\left(x/\theta\right)^{1/4}}=e^{-\left(x/\theta\right)^{1/4}}\frac{1}{4\theta}\left(x/\theta\right)^{-3/4} \quad \text{(not hard)}$$

$$E(X)=\int_{0}^{+\infty}\Pr(X>x)dx=\int_{0}^{+\infty}e^{-\left(x/\theta\right)^{1/4}}dx$$

Don't know what to do next? Simply do a transformation.

Let set $\left(\frac{x}{\theta}\right)^{1/4}=t$. This will change the ugly $e^{-\left(\frac{x}{\theta}\right)^{1/4}}$ into a nice e^{-t}.

$$\left(\frac{x}{\theta}\right)^{1/4}=t \Rightarrow \frac{x}{\theta}=t^4 \Rightarrow x=\theta t^4 \quad \Rightarrow dx=4\theta t^3 dt$$

$$E(X)=\int_0^{+\infty} e^{-\left(\frac{x}{\theta}\right)^{1/4}}dx=\int_0^{+\infty} e^{-t}4\theta t^3 dt=4\theta\int_0^{+\infty} e^{-t}t^3 dt=4\theta(3!)=(4!)\theta=240$$

Please note that we use the formula $\int_0^{+\infty} e^{-t}t^n dt=n!$ in the chapter on gamma distribution.

Similarly, we can find $E(X^2)$.

$$E(X^2)=\int_0^{+\infty}\Pr(X>x)dx^2=\int_0^{+\infty} e^{-\left(\frac{x}{\theta}\right)^{1/4}}dx^2=\int_0^{+\infty} e^{-\left(\frac{x}{\theta}\right)^{1/4}}2x dx$$
$$=\int_0^{+\infty} e^{-t}2\left(\theta t^4\right)\left(4\theta t^3 dt\right)=8\theta^2\int_0^{+\infty} e^{-t}t^7 dt=8\theta^2(7!)=(8!)\theta^2=4,032,000$$

$$Var(X)=E(X^2)-E^2(X)=(8!)\theta^2-\left[(4!)\theta\right]^2=3,974,400$$

Next, we will find the 80th percentile, x_{80}.

$$F(x_{80})=80\% \quad \Rightarrow \Pr(X>x_{80})=20\%$$

$$e^{-\left(\frac{x_{80}}{10}\right)^{1/4}}=20\% \quad \Rightarrow -\left(\frac{x_{80}}{10}\right)^{1/4}=\ln 20\% \quad \Rightarrow x_{80}=10(-\ln 20\%)^4\approx 67.1$$

$$\Pr(X>20)=e^{-\left(\frac{20}{10}\right)^{1/4}}=.3045$$

Method 1 – Use memorized formulas (unpleasant to memorize, but you will work faster in the exam if you have them memorized).

Remember, under Method 1, we still do not want to memorize the pdf. We will derive the pdf instead.

Let's start with $\Pr(X>x)$.

$$\Pr(X > x) = e^{-(x/\theta)^{\beta}} = e^{-\left(x/\theta\right)^{1/4}} \quad \text{where } x \geq 0$$

$$f(x) = -\frac{d}{dx}\Pr(X > x) = -\frac{d}{dx}e^{-\left(x/\theta\right)^{1/4}} = e^{-\left(x/\theta\right)^{1/4}} \frac{1}{4\theta}\left(x/\theta\right)^{-3/4} \quad \text{(not hard)}$$

To find the mean and variance, we use the memorized formula:

$$E(X^n) = \theta^n\,\Gamma\left(1+\frac{n}{\beta}\right)$$

We have

$$E(X) = \theta\,\Gamma\left(1+\frac{1}{\beta}\right) = \theta\,\Gamma\left(1+\frac{1}{1/4}\right) = \theta\,\Gamma(5) = (4!)\theta$$

$$E(X^2) = \theta^2\,\Gamma\left(1+\frac{2}{\beta}\right) = \theta^2\,\Gamma\left(1+\frac{2}{1/4}\right) = \theta^2\,\Gamma(9) = (8!)\theta^2$$

$$Var(X) = E(X^2) - E^2(X) = \ldots$$

You see that we get the same result as Method 1.

The calculation of x_{80} and $\Pr(X > 20)$ is the same as the calculation in Method 2.

Homework for you: Rework all the problems in this chapter.

Chapter 23 Pareto distribution

Key points:

1. Pareto distribution is often used to model losses where the probability of incurring a catastrophic loss (such as in a hurricane) is small but not zero.

2. You need to memorize two formulas:

$\Pr(X > x) = \left(\dfrac{\beta}{x+\beta}\right)^{\alpha}$, where α, β are positive numbers and $0 \le x < +\infty$.

for $k < \alpha$, the k-th moment is

$$E\left(X^k\right) = \frac{\beta^k}{C_{\alpha-1}^k}, \quad \text{where } C_{\alpha-1}^k = \frac{(\alpha-1)!}{k!\left[(\alpha-1)-k\right]!} = \frac{(\alpha-1)(\alpha-2)...(\alpha-k)}{k!}$$

Please note that in the k-th moment formula, α is not necessarily an integer. When α is a positive non-integer, we still use the notation $C_{\alpha-1}^k$ to help us memorize the formula.

However, if $k \ge \alpha$, then $E\left(X^k\right)$ is not defined (we will look into this with an example).

Explanations

Let's use both Pareto distribution and exponential distribution to model a loss random variable X.

Exponential: $\Pr\left(X > x\right) = e^{-x/\theta}$

Pareto $\Pr(X > x) = \left(\dfrac{\beta}{x+\beta}\right)^{\alpha}$

In exponential distribution, the probability that loss exceeds a large threshold amount (such as \$1,000,000) is close to zero. You can see, mathematically, that when X gets bigger, $\Pr\left(X > x\right)$ quickly approaches zero.

For example, if we set $\theta = \$5,000$ (average loss amount), then the probability that loss exceeds \$1,000,000 is:

$$\Pr\left(X > x\right) = e^{-x/\theta} = e^{-1,000,000/5,000} = e^{-200} = 1.384 \times 10^{-87} \approx 0$$

In contrast, in Pareto distribution, the probability that loss exceeds a large threshold amount, though small, is not zero.

For example, let's set $\beta = \$5,000$, $\alpha = 2$. The probability that loss exceeds \$1,000,000 is:

$$\Pr(X > x) = \left(\frac{\beta}{x+\beta}\right)^{\alpha} = \left(\frac{5,000}{1,000,000+5,000}\right)^{2} = 0.004975^2 = 0.00002475$$

0.00002475 is small, but is significantly more than 1.384×10^{-87}.

If you are an actuary trying to fit a bunch of loss data into a neat formula, you need to ask yourself, "How likely is a catastrophic loss? Is the probability so small that I can ignore it?" If "No," then you can use Pareto distribution to fit the loss data. If "Yes," then you can use exponential (or some other models).

This brings up an important concept called a "tail." For two probability distributions A and B (such as exponential and Pareto) and for an identical loss amount x, if $\Pr^A(X > x) > \Pr^B(X > x)$, we say that A has a fatter (or heavier) tail than does B. For example, Pareto distribution has a fatter or heavier tail than does exponential distribution.

Stated differently, if a probability distribution has a fat (or heavy) tail, then the probability that loss exceeds a large threshold amount approaches zero slowly. In contrast, if a probability distribution has a light tail, then the probability that loss exceeds a large threshold amount approaches zero quickly.

Pareto, gamma, and lognormal distributions have a heavy tail. Exponential has a light tail. Heavy-tailed distributions are better at modeling losses where there's a small but non-zero chance of having a large loss.

Now let's turn our attention to the *k*-th moment formula:

$$E\left(X^k\right) = \frac{\beta^k}{C_{\alpha-1}^k}$$

The derivation of this formula is time-consuming. Just memorize the above formula.

However, let's take time to derive the formula for Pareto mean and variance. This way, if you forget the formula in the exam, you can still calculate the mean and variance using basic statistics principles.

To derive the mean and variance formula, we will use the following equation:

$$E(X^n) = \int_0^{+\infty} \Pr(X > x) dx^n = \int_0^{+\infty} \left(\frac{\beta}{x+\beta}\right)^{\alpha} dx^n$$

$$E(X) = \int_0^{+\infty} \Pr(X > x) dx = \int_0^{+\infty} \left(\frac{\beta}{x+\beta}\right)^{\alpha} dx = \int_{\beta}^{+\infty} \left(\frac{\beta}{t}\right)^{\alpha} dt \quad (\text{let } x+\beta = t)$$

$$= \beta^{\alpha} \int_{\theta}^{+\infty} t^{-\alpha} dt$$

If $-\alpha + 1 < 0$ (i.e. $\alpha > 1$), then

$$E(X) = \beta^{\alpha} \int_{\beta}^{+\infty} t^{-\alpha} dt = \beta^{\alpha} \left[\frac{t^{-\alpha+1}}{-\alpha+1}\right]_{\beta}^{+\infty} = \beta^{\alpha} \frac{1}{-\alpha+1}\left[+\infty^{-\alpha+1} - \beta^{-\alpha+1}\right] = \frac{\beta}{\alpha - 1}$$

If $\alpha = 1$, $E(X) = \beta \int_{\beta}^{+\infty} t^{-1} dt = \beta[\ln t]_{\beta}^{+\infty}$ (undefined)

If $\alpha < 1$

$$E(X) = \beta^{\alpha} \int_{\beta}^{+\infty} t^{-\alpha} dt = \beta^{\alpha} \left[\frac{t^{-\alpha+1}}{-\alpha+1}\right]_{\beta}^{+\infty} = \beta^{\alpha} \frac{1}{-\alpha+1}\left[+\infty^{1-\alpha} - \beta^{-\alpha+1}\right] \text{ (undefined)}$$

$$E(X^2) = \int_0^{+\infty} \Pr(X > x) dx^2 = \int_0^{+\infty} 2x \left(\frac{\beta}{x+\beta}\right)^{\alpha} dx = 2\int_{\beta}^{+\infty} (t-\beta)\left(\frac{\beta}{t}\right)^{\alpha} dt$$

$$= 2\left[\int_{\beta}^{+\infty} t\left(\frac{\beta}{t}\right)^{\alpha} dt - \int_{\beta}^{+\infty} \beta\left(\frac{\beta}{t}\right)^{\alpha} dt\right]$$

$$\int_{\beta}^{+\infty} t\left(\frac{\beta}{t}\right)^{\alpha} dt = \beta^{\alpha} \int_{\beta}^{+\infty} t^{1-\alpha} dt = \beta^{\alpha}\left[\frac{t^{1-\alpha+1}}{1-\alpha+1}\right]_{\beta}^{+\infty} = \beta^{\alpha}\left[\frac{t^{2-\alpha}}{2-\alpha}\right]_{\beta}^{+\infty}$$

If $\alpha > 2$, then

$$\beta^{\alpha}\left[\frac{t^{2-\alpha}}{2-\alpha}\right]_{\beta}^{+\infty} = \beta^{\alpha}\frac{1}{2-\alpha}\left[+\infty^{2-\alpha} - \beta^{2-\alpha}\right] = \frac{\beta^2}{\alpha-2}$$

$$\int_{\beta}^{+\infty} \beta\left(\frac{\beta}{t}\right)^{\alpha} dt = \beta^{\alpha+1}\int_{\beta}^{+\infty} t^{-\alpha} dt = \beta^{\alpha+1}\left[\frac{t^{-\alpha+1}}{-\alpha+1}\right]_{\beta}^{+\infty} = \beta^{\alpha+1}\frac{1}{-\alpha+1}\left[+\infty^{-\alpha+1} - \beta^{-\alpha+1}\right]$$

If $\alpha > 1$

$$\beta^{\alpha+1}\frac{1}{-\alpha+1}\left[+\infty^{-\alpha+1}-\beta^{-\alpha+1}\right]=\beta^{\alpha+1}\frac{-\beta^{-\alpha+1}}{-\alpha+1}=\frac{\beta^2}{\alpha-1}$$

Finally for $\alpha > 2$:

$$E\left(X^2\right)=2\left[\int_{\beta}^{+\infty}t\left(\frac{\beta}{t}\right)^{\alpha}dt-\int_{\beta}^{+\infty}\beta\left(\frac{\beta}{t}\right)^{\alpha}dt\right]=2\left[\frac{\beta^2}{\alpha-2}-\frac{\beta^2}{\alpha-1}\right]=\frac{2\beta^2}{(\alpha-1)(\alpha-2)}=\frac{\beta^2}{C_{\alpha-1}^2}$$

You can easily verify that $E\left(X^2\right)$ is undefined if $\alpha \le 2$.

Sample problems and solutions

Problem 1

A Pareto distribution has the parameters $\beta = 100$ and $\alpha = 2$. Find the 60th percentile.

Solution

$$\Pr(X > x)=\left(\frac{\beta}{x+\beta}\right)^{\alpha}=\left(\frac{100}{x+100}\right)^{2}$$

Let x_{60} represent the 60th percentile.

$$F(x_{60})=60\% \quad \Rightarrow \Pr(X > x_{60})=40\% \quad \Rightarrow \left(\frac{100}{x_{60}+100}\right)^2=40\% \quad \Rightarrow x_{60}=58.11$$

Problem 2

Individual claims on an insurance policy have a Pareto distribution with a mean of \$3,000 and a variance of $\$^2$ 63,000,000 (the unit is dollar squared). This policy has a deductible of \$5,000. The maximum amount the insurer will pay on an individual claim is \$9,000.

Find
(1) The expected claim payment the insurer will pay on this policy.
(2) The expected claim payment the insurer will pay on this policy, given
there is a claim.

Solution

Let X represent the individual loss; let Y represent the claim payment by the insurance company. We are asked to find $E(Y)$ and $E(Y|Y>0)$.

We first need to find the two parameters of the Pareto distribution.

$$E\left(X^k\right)=\frac{\beta^k}{C_{\alpha-1}^k}$$
$$\Rightarrow \quad E(X)=\frac{\beta}{\alpha-1},\ E\left(X^2\right)=\frac{\beta^2}{C_{\alpha-1}^2}=\frac{2\beta^2}{(\alpha-1)(\alpha-2)}$$
$$\Rightarrow \quad Var(X)=E\left(X^2\right)-E^2(X)=\frac{2\beta^2}{(\alpha-1)(\alpha-2)}-\left(\frac{\beta}{\alpha-1}\right)^2=\left(\frac{\beta}{\alpha-1}\right)^2\frac{\alpha}{\alpha-2}$$

We are given $E(X)=3K$, $Var(X)=63K^2$, where K=\$1,000.

$$\frac{\beta}{\alpha-1}=3K,\ \left(\frac{\beta}{\alpha-1}\right)^2\frac{\alpha}{\alpha-2}=63K^2 \quad \Rightarrow \alpha=\frac{7}{3},\beta=4K$$

We use K=\$1,000 to keep our calculation simple and fast.

Then $\Pr(X>x)=\left(\frac{\beta}{x+\beta}\right)^{\alpha}=\left(\frac{4}{x+4}\right)^{\frac{7}{3}}$

x is the loss amount in thousands of dollars

$$Y=\begin{cases}0, & \text{if } X\le 5K\\ (X-5)K, & \text{if } 5K<X\le 14K\\ 9K, & \text{if } X>14K\end{cases}$$

The expected claim payment:

$$E(Y)=\int_5^{14}(x-5)f(x)dx+\int_{14}^{+\infty}9f(x)dx$$

Since we don't know $f(x)$, we will solve the above integration using $\Pr(X>x)$.

$$f(x)=\frac{d\left[F(x)\right]}{dx} \quad \Rightarrow f(x)\,dx=d\left[F(x)\right]=d\left[1-\Pr(X>x)\right]=-d\left[\Pr(X>x)\right]$$

$$\int_5^{14}(X-5)f(x)\,dx=-\int_5^{14}(x-5)d\left[\Pr(X>x)\right]$$

$$=\int_5^{14}\Pr(X>x)d\left[(x-5)\right]-\left[(x-5)\Pr(X>x)\right]_5^{14} \quad \text{(integration by parts)}$$

$$=\int_5^{14}\Pr(X>x)dx-(14-5)\Pr(X>14)+(5-5)\Pr(X>5)$$

$$=\int_5^{14}\Pr(X>x)dx-9\Pr(X>14)$$

$$\int_{14}^{+\infty}9f(x)\,dx=9\int_{14}^{+\infty}f(x)dx=9\left[\Pr(X>14)\right]$$

$$\Rightarrow \quad E(Y)=\int_5^{14}(x-5)f(x)\,dx+\int_{14}^{+\infty}9f(x)\,dx=\int_5^{14}\Pr(X>x)dx$$

Generally, for a random loss variable X, if there is a deductible of $d\geq 0$ and a maximum payment of $L\geq 0$, then the expected payment is

$$E(Y)=\int_d^{d+L}\Pr(X>x)\,dx$$

If there is no deductible and no limit on how much the insurer will pay (i.e. d =0 and $L=+\infty$), then the expected payment by the insurer is:

$$E(Y)=\int_0^{+\infty}\Pr(X>x)dx \quad \text{(we looked at this formula in Chapter 20)}$$

Now we are ready to find the expected payment:

$$\Rightarrow \quad E(Y)=\int_5^{14}\Pr(X>x)dx=\int_5^{14}\left(\frac{4}{x+4}\right)^{\frac{7}{3}}dx=\int_9^{18}\left(\frac{4}{t}\right)^{\frac{7}{3}}dt=4^{\frac{7}{3}}\int_9^{18}t^{-\frac{7}{3}}dt$$

$$=4^{\frac{7}{3}}\frac{1}{-\frac{7}{3}+1}\left[t^{-\frac{7}{3}+1}\right]_9^{18}=0.61372K=\$613.72$$

The expected claim payment the insurer will pay on this policy, given that there is a claim:

$$E\left(Y \middle| X > 5\right) = \frac{E(Y)}{\Pr(Y > 0)} = \frac{E(Y)}{\Pr(X > 5)}$$

$$\Pr(X > 5) = \left(\frac{4}{5+4}\right)^{\frac{7}{3}} = \left(\frac{4}{9}\right)^{\frac{7}{3}}$$

$$\Rightarrow \quad E\left(Y \middle| X > 5\right) = \frac{E(Y)}{\Pr(X > 5)} = \$613.72\left(\frac{4}{9}\right)^{-\frac{7}{3}} = \$4,071.26$$

Homework for you: Rework all the problems in this chapter.

Chapter 24 Normal distribution

Most students are pretty comfortable with normal distribution. So I will give a quick outline.

1. Why normal distribution is important

- The sample mean is approximately normally distributed, no matter what form the population variable is distributed in (Central limit theorem).

  If a random sample of size n is taken from a population that has a mean $E(X)$ and a standard deviation σ, then the sample mean $\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ is approximately normally distributed with a mean of $E(X)$ and a standard deviation of $\frac{\sigma}{\sqrt{n}}$ for large *n*. So $\bar{X} \sim N\left[E(X), \frac{\sigma}{\sqrt{n}}\right]$ for large *n*.

- The sum of a large number of independent identically distributed random variables is approximately normally distributed. (another version of the Central limit theorem)

  If $X_1, X_2, \ldots, X_n$ are independent identically distributed with a mean of $E(X)$ and a standard deviation σ, then $S = X_1 + X_2 + \ldots + X_n$ is approximately normally distributed with mean $nE(X)$ and standard deviation $\sqrt{n}\sigma$ (regardless of what kind of distribution $X_1, X_2, \ldots, X_n$ have). That is to say $\sum_{i=1}^{n} X_i$ is $N\left[nE(X), \sqrt{n}\sigma\right]$.

- A rule of thumb: If a large number of small effects are acting additively, normal distribution can probably be assumed (such as the sum of independent identically distributed loss random variables)

  In contrast, if a large number of small effects are acting multiplicatively, normal distribution should NOT be assumed. Instead, a lognormal distribution can probably be assumed (for example, compound interest rates have multiplicative effects and can be modeled with lognormal distribution).

2. The probability density function (pdf) has two parameters: mean $E(X)$ and standard deviation σ. A normal random variable $X \sim N\left[E(X), \sigma\right]$ has the mean $E(X)$ and the standard deviation σ.

Normal distribution's pdf and cdf are complex and have never been directly tested in SOA exams. So there is no need to memorize the formula:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left\{\left[x - E(X)\right]/\sigma\right\}^2} \quad (-\infty < x < +\infty)$$

3. Standard normal distribution $Z \sim N(0,1)$ is obtained by transforming normal distribution $X \sim N\left[E(X), \sigma\right]$ using $Z = \dfrac{x - E(X)}{\sigma}$.

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (-\infty < z < +\infty)$$

This pdf is a special case of a normal distribution's pdf where $E(X) = 0$ and $\sigma = 1$.

4. The accumulative density function for standard normal distribution $Z \sim N(0,1)$:

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-w^2/2} dw = \Phi(z)$$

For a $z \le 0$, you'll need the equation $\Phi(z) + \Phi(-z) = 1$. The equation $\Phi(z) + \Phi(-z) = 1$ stands true for any $-\infty \le z \le +\infty$.

5. A linear combination of two or more independent normal random variables is normal. If $X \sim N\left(E(X), \sigma_X\right)$, $Y \sim N\left(E(Y), \sigma_y\right)$, and X and Y are independent, then

$$Z = aX + bY + c \sim N\left[aE(X) + bE(Y) + c, \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2}\right]$$

6. Continuity correction factor

Binomial distribution $B(X, n, p)$ approaches $N(np, \sqrt{np(1-p)})$ for large *n* and *p* not too close to 0 or 1. Poisson distribution $P(X, \lambda)$ approaches normal distribution $N(\lambda, \sqrt{\lambda})$

To use normal approximation to find the probability of having x occurrences (in binomial distribution or Poisson distribution), remember to add or subtract the continuity correction factor.

Let X represent the number of successes, while Y represents the normal random variable used to approximate X, then

$$\Pr(X = k) = \Pr(k - 0.5 < Y < k + 0.5) = \Pr(k - 0.5 \le Y < k + 0.5)$$

$$= \Phi\left(\frac{k + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{k - 0.5 - \mu}{\sigma}\right)$$

Since Y is continuous, we have $\Pr(Y \ge a) = \Pr(Y > a)$, $\Pr(Y \le a) = \Pr(Y < a)$.

So $X = k$ corresponds to $Y \in (k - 0.5, k + 0.5)$. A single point k becomes a range $(k - 0.5, k + 0.5)$.

$$\Pr(X < k) = \Pr(Y < k - 0.5) = \Pr(Y \le k - 0.5) = \Phi\left(\frac{k - 0.5 - \mu}{\sigma}\right)$$

$$\Pr(X > k) = \Pr(Y > k + 0.5) = 1 - \Phi\left(\frac{k - 0.5 - \mu}{\sigma}\right)$$

$$\Pr(X \le k) = \Pr(X < k + 1) = \Pr(Y < k + 1 - 0.5)$$
$$= \Pr(Y < k + 0.5) = \Phi\left(\frac{k + 0.5 - \mu}{\sigma}\right)$$

$$\Pr(X \ge k) = \Pr(X > k - 1) = \Pr(Y > k - 1 + 0.5)$$
$$= \Pr(Y > k - 0.5) = 1 - = \Phi\left(\frac{k - 0.5 - \mu}{\sigma}\right)$$

Sample Problems and Solutions

Problem 1

The claim sizes for a particular type of policy are normally distributed with a mean of \$5,000 and a standard deviation of \$500. Determine the probability that two randomly chosen claims differ by \$600.

Solution

Let X_1, X_2 represent two claim sizes randomly chosen. X_1, X_2 are normally distributed with a mean of 5,000 and a standard deviation of 500.

Let $Y = X_1 - X_2$. Then Y is normally distributed (the linear combination of normal random variables is also normal). We are asked to find $\Pr(|Y| > 600)$.

$$E(Y) = E(X_1 - X_2) = E(X_1) - E(X_2) = 0$$

Because X_1, X_2 are independent, we have

$$Var(Y) = Var(X_1 - X_2) = Var(X_1) + Var(X_2) = 2 \times 500^2$$

$$\Pr(|Y| > 600) = \Pr\left(|z| > \frac{600 - \mu_Y}{\sigma_Y}\right) = \Pr\left(|z| > \frac{600 - 0}{500\sqrt{2}}\right) = \Pr(|z| > 0.8485)$$

$$\Pr(|z| > 0.8485) = 1 - \Pr(|z| < 0.8485) = 1 - \Pr(-0.8485 < z < 0.8485)$$

$$\begin{aligned}\Pr(-0.8485 < z < 0.8485) &= \Phi(0.8485) - \Phi(-0.8485) \\ &= \Phi(0.8485) - [1 - \Phi(0.8485)] \\ &= 2\Phi(0.8485) - 1 = 0.6038\end{aligned}$$

$$\Pr(|z| > 0.8485) = 1 - \Pr(|z| < 0.8485) = 0.3962$$

$$\Pr(|Y| > 600) = \Pr(|z| > 0.8485) = 0.3962$$

Problem 2

The annual claim amount for an auto insurance policy has a mean of $30,000 and a standard deviation of $5,000. A block of auto insurance has 25 such policies.

Assume individual claims from this block of auto insurance are independent. Using normal approximation, find the probability that the aggregate annual claims for this block of auto insurance exceed $800,000.

Solution

Let $X(i)$=annual claims amount for the *i*-th auto insurance policy (in $1,000), Y =the aggregate annual claims amount for the block of auto insurance (in $1,000).

$X(1), X(2), \ldots, X(25)$ are independent identically distributed with a mean of 30 and a standard deviation of 5.

$Y = X(1) + X(2) + \ldots + X(25)$ is approximately normal.

$E(Y) = 25E(X(i)) = 25(30) = 750$

$Var(Y) = 25Var(X(i)) = 25(5)^2$, $\sigma_Y = \sqrt{Var(Y)} = 25$

$$z = \frac{Y - E(Y)}{\sigma_Y} = \frac{800 - 750}{25} = 2$$

$\Pr(Y < 800) = \Phi(2) = 97.72\%$

$\Pr(Y > 800) = 1 - \Pr(Y < 800) = 1 - 97.72\% = 2.28\%$

So there is a 2.28% chance that the aggregate annual claim amount for the 25 insurance policies exceeds $800,000.

Homework for you: #9,#19 May 2000; #6, Nov 2000; #19, May 2001; #15, #40 Nov 2001; #13, May 2003.

Chapter 25 Lognormal distribution

If a positive random variable Y $(Y>0)$ is the product of n independent identically distributed random variables $X_1, X_2, \ldots, X_n$

$$Y = X_1 X_2 \ldots X_n$$

Then

$$\ln Y = \ln X_1 + \ln X_2 + \ldots + \ln X_n$$

Because $\ln X_1, \ln X_2, \ldots \ln X_n$ are independent identically distributed, $\ln Y$ is approximately normally distributed (Central limit theorem).

If the normal random variable $\ln Y$ has a mean of $\mu_{\ln Y}$ and a standard deviation of $\sigma_{\ln Y}$, then

pdf -- no need to memorize: $f(y) = \dfrac{1}{y} \underbrace{\dfrac{1}{\sqrt{2\pi}\sigma_{\ln Y}} e^{-\frac{\left[\ln y - \mu_{\ln Y}\right]^2}{2\sigma_{\ln Y}^2}}}_{\text{normal pdf for } \ln Y}$

cdf -- most important formula that you must memorize:

$$F(y) = \Pr(Y \le y) = \Pr(\ln Y \le \ln y) = \Phi\left[\frac{\ln y - \mu_{\ln Y}}{\sigma_{\ln Y}}\right]$$

Mean – good if you can memorize: $E(Y) = e^{\mu_{\ln Y} + \frac{1}{2}\sigma_{\ln Y}^2}$

Variance -- good if you can memorize:

$$Var(Y) = \left(e^{\sigma_{\ln Y}^2} - 1\right) e^{2\mu_{\ln Y} + \sigma_{\ln Y}^2} = \left(e^{\sigma_{\ln Y}^2} - 1\right) E^2(Y)$$

Please note in the above formulas, $\mu_{\ln Y}$ and $\sigma_{\ln Y}$ are the mean and standard deviation of $\ln Y$, not the mean and standard deviation of Y.

In the heat of the exam, it is very easy for candidates to mistakenly use the mean and standard deviation of Y as the two parameters for the lognormal random variable Y. To help avoid this kind of mistake, use $\mu_{\ln Y}$ and $\sigma_{\ln Y}$ instead of the standard notations μ and σ.

Let's look at the pdf formula. $\ln Y$ is normally distributed with mean $\mu_{\ln Y}$ and standard deviation $\sigma_{\ln Y}$. Applying the normal pdf formula (see Chapter 22), we get the pdf for $\ln Y$:

$$f(\ln y) = \frac{1}{\sqrt{2\pi}\sigma_{\ln Y}} e^{-\frac{\left[\ln y - \mu_{\ln Y}\right]^2}{2\sigma_{\ln Y}{}^2}} \qquad (-\infty < \ln y < +\infty)$$

To find the pdf for Y, notice $\frac{d}{dy}\ln y = \frac{1}{y} > 0$. So $\ln Y$ is an increasing function.

$$\Rightarrow \quad F(y) = \Pr(Y \le y) = \Pr(\ln Y \le \ln y) = F(\ln y)$$

$$\Rightarrow \quad f(y) = \frac{d}{dy}F(y) = \frac{d}{dy}F(\ln y) = \frac{dF(\ln y)}{d(\ln y)}\frac{d(\ln y)}{dy} = f(\ln y)\frac{1}{y}$$

$$\Rightarrow \quad f(y) = \frac{1}{y}\underbrace{\frac{1}{\sqrt{2\pi}\sigma_{\ln Y}} e^{-\frac{\left[\ln y - \mu_{\ln Y}\right]^2}{2\sigma_{\ln Y}{}^2}}}_{\substack{\text{normal pdf} \\ \text{for } \ln Y}}$$

The cdf formula is self-explanatory. It is the most important lognormal formula and is very likely to be tested on Exam P.

The mean formula is counter-intuitive. We might expect that $E(Y) = e^{E(\ln Y)} = e^{\mu_{\ln Y}}$.

However, the correct formula is

$$E(Y) = e^{\mu_{\ln Y} + \frac{1}{2}\sigma_{\ln Y}{}^2}.$$

The mathematical proof for the mean formula is complex. To get an intuitive feel for this formula without using a rigorous proof, notice that the pdf for $\ln Y$ is NOT a bell curve ---the pdf for $\ln Y$ is skewed to the right (you can look at a pdf graph from a textbook).

As a result, there is a factor of $e^{\frac{1}{2}\sigma_{\ln Y}{}^2}$ applied to $e^{\mu_{\ln Y}}$:

$$E(Y) = e^{\mu_{\ln Y} + \frac{1}{2}\sigma_{\ln Y}{}^2} = e^{\mu_{\ln Y}} e^{\frac{1}{2}\sigma_{\ln Y}{}^2}$$

It is hard to find an intuitive explanation for the variance formula. To be safe, you might want to memorize it.

Another point. When solving problems, many candidates find it difficult to determine whether a random variable is normally distributed or lognormally distributed. To avoid the confusion, remember the follow two rules:

Rule 1

If X is normal with parameters of mean μ and standard deviation σ, then e^X is lognormal with parameters of mean μ and standard deviation σ.

Rule 2

If Y is lognormal with parameters of mean μ and standard deviation σ, then $\ln Y$ is normal with parameters of mean μ and standard deviation σ.

These two rules will come in handy when you solve a difficult problem.

Another point. The product of independent lognormal random variables is also lognormal.

If X is lognormal $\left(\mu_{\ln X}, \sigma_{\ln X}\right)$, Y is lognormal $\left(\mu_{\ln Y}, \sigma_{\ln Y}\right)$, X, Y are independent, then XY is also lognormal with the following parameters:

$$E\left[\ln(XY)\right] = E(\ln X + \ln Y) = E(\ln X) + E(\ln Y) = \mu_{\ln X} + \mu_{\ln Y}$$

$$Var\left[\ln(XY)\right] = Var(\ln X + \ln Y) = Var(\ln X) + Var(\ln Y) = \sigma_{\ln X}{}^2 + \sigma_{\ln Y}{}^2$$

$$\Rightarrow \quad XY \text{ is lognormal}\left(\mu_{\ln X} + \mu_{\ln Y}, \sqrt{\sigma_{\ln X}{}^2 + \sigma_{\ln Y}{}^2}\right)$$

Let's see why. If X and Y are lognormal, then $\ln X$ and $\ln Y$ are normal (Rule 2). The sum of two independent normal random variables is also normal. So $\ln X + \ln Y = \ln(XY)$ is normal.

Finally, for a lognormal random variable Y, if you are given the mean $E(Y)$ and variance $Var(Y)$, then you will need to find the lognormal parameters by solving the following equations (SOA can easily write a question like this):

$$E(Y) = e^{\mu_{\ln Y} + \frac{1}{2}\sigma_{\ln Y}^2}$$

$$Var(Y) = \left(e^{\sigma_{\ln Y}^2} - 1\right) E^2(Y)$$

$$\Rightarrow \quad \sigma_{\ln Y} = \sqrt{\ln\left[1 + \frac{Var(Y)}{E^2(Y)}\right]}, \quad \mu_{\ln Y} = \ln E(Y) - \frac{1}{2}\sigma_{\ln Y}^2$$

Caution: Don't write $\sigma_{\ln Y} = \ln\sqrt{\left[1 + \frac{Var(Y)}{E^2(Y)}\right]}$. This is a common mistake. Make sure you write $\sigma_{\ln Y} = \sqrt{\ln(...)}$, not $\sigma_{\ln Y} = \ln\sqrt{(...)}$.

Lognormal distribution has wide applications in insurance and other fields. Generally, the product of many independent small effects is likely to be lognormally distributed. Lognormal distribution is a multiplicative process. In contrast, normal distribution is an additive process (the sum of independent identically distributed random variables is approximately normal).

For example, the number of defects in software is a multiplicative (or compounding) process and thus approximately lognormal. The market value of an asset is the compounding effect of an interest rate and is approximately lognormal.

To see why the number of defects or mistakes is a compounding process, think of a simple example. Compare two outcomes – you pass Exam P or fail Exam P. If you fail Exam P, the financial effect is not simply the sum of the wasted exam fee and lost income you could have earned while you were studying for Exam P. The effect is far more. Failing Exam P will delay your job search, which in turn will delay your promotion, which in turn will cause other effects. Defects or mistakes have a compounding process.

If you forget everything else about lognormal distribution, remember this: if X is lognormal, then $\ln X$ is normal.

Sample Problems and Solutions

Problem 1

The value of investing \$1 in stocks for 20 years is G (dollars). Assume that Y is lognormal with the parameters $\mu_{\ln Y} = 2$ and $\sigma_{\ln Y} = 0.3$.

Find
(1) The probability that Y is between 6 and 8;
(2) $E(Y)$ and $Var(Y)$.

Solution

$$F(y) = P(Y \le y) = P(\ln Y \le \ln y) = \Phi\left[\frac{\ln y - \mu_{\ln Y}}{\sigma_{\ln Y}}\right] = \Phi\left(\frac{\ln y - 2}{0.3}\right)$$

$$Pr(6 < Y < 8) = \Phi\left(\frac{\ln 8 - 2}{0.3}\right) - \Phi\left(\frac{\ln 6 - 2}{0.3}\right)$$

$$\Phi\left(\frac{\ln 6 - 2}{0.3}\right) = \Phi(-0.6941) = 1 - \Phi(0.6941) \approx 0.24$$

$$\Phi\left(\frac{\ln 8 - 2}{0.3}\right) = \Phi(0.2648) \approx 0.60$$

$$Pr(6 < Y < 8) \approx 0.60 - 0.24 = 36\%$$

Please note that $Pr(6 < Y < 8) = Pr(6 \le Y \le 8) = Pr(6 < Y \le 8) = Pr(6 \le Y < 8)$ because Y is continuous.

$$E(Y) = e^{\mu_{\ln Y} + \frac{1}{2}\sigma_{\ln Y}^2} = e^{2 + \frac{1}{2}0.3^2} = 7.729$$

$$Var(Y) = \left(e^{\sigma_{\ln Y}^2} - 1\right)E^2(Y) = Var(Y) = \left(e^{0.3^2} - 1\right)(7.729)^2 = 5.626$$

Problem 2

You are given the 20th and 80th percentile of individual claims X :

(1) the 20th percentile is 61.04
(2) the 80th percentile is 85.49

Given X is lognormally distributed. Find the probability that a claim exceeds 92.

Solution

$$\frac{d}{dx}\ln x = \frac{1}{x} > 0 \quad \text{(for any } x > 0\text{)}$$

So $Y = \ln x$ is an increasing function. Then the 20th and 80th percentile of X corresponds to the 20th and 80th percentile of $Y = \ln x$.

We can easily prove this. Let $x_{0.2}$ and $x_{0.8}$ represent the 20% and 80% percentile of X . Let $y_{0.2}$ and $y_{0.8}$ represent the 20% and 80% percentile of y .

$$0.2 = \Pr(X \le x_{0.2}) = \Pr(\ln X \le \ln x_{0.2}) \quad \Rightarrow y_{0.2} = \ln x_{0.2}$$
$$0.8 = \Pr(X \le x_{0.8}) = \Pr(\ln X \le \ln x_{0.8}) \quad \Rightarrow y_{0.8} = \ln x_{0.8}$$

From the normal table, we find that the z for 0.8 is 0.842. In other words, $\Phi(0.842) = 0.8$. Then $\Phi(-0.842) = 1 - \Phi(0.842) = 0.2$.

Using the formula (Y is normal)

$$F(x) = \Phi\left[\frac{\ln x - \mu}{\sigma}\right]$$

we have $\quad \dfrac{\ln 61.04 - \mu}{\sigma} = -0.842, \quad \dfrac{\ln 85.49 - \mu}{\sigma} = 0.842$

Solve the equations:

$$\mu = 4.28, \quad \sigma = 0.2$$

Then $\Pr(X > 92) = 1 - \Phi\left(\dfrac{\ln 92 - 4.28}{0.2}\right) = 1 - \Phi(1.21) = 1 - .89 = 0.11$

Problem 3

An actuary models losses due to large fires using a lognormal distribution. The average loss due to a large fire is $30 million. The standard deviation of losses due to large fires is $10 million.

Calculate the probability that a loss due to a large fire exceeds $35 million.

Solution

Let X represent the loss amount (in million dollars) in a large fire. We are given the following information:

- X is lognormal
- $E(X)=30$
- $\sigma(X)=10$

We are asked to find $P(X>35)$.

First, we need to solve for the two parameters of the lognormal random variable X :

$$\begin{cases} E(X)=e^{\mu_{\ln X}+\frac{1}{2}\sigma_{\ln X}{}^{2}}=30 \\ Var(X)=\left(e^{\sigma_{\ln X}{}^{2}}-1\right)E^{2}(X)=10^{2} \end{cases}$$

$$\Rightarrow \quad \sigma_{\ln X}=\sqrt{\ln\left[1+\frac{Var(X)}{E^{2}(X)}\right]}=\sqrt{\ln\left[1+\left(\frac{10}{30}\right)^{2}\right]}=0.3246$$

$$\mu_{\ln X}=\ln E(X)-\frac{1}{2}\sigma_{\ln X}{}^{2}=\ln 30-\frac{1}{2}(0.3246)^{2}=3.3485$$

$$P(X>35)=P(\ln X>\ln 35)=1-P(\ln X\leq \ln 35)=1-\Phi\left(\frac{\ln 35-3.3485}{0.3246}\right)$$

$$=1-\Phi(0.637)=1-0.7389\approx 0.26$$

Problem 4

The cumulative value of investing \$1 for 20 years, Y, is calculated as follows:

$$Y=(1+i_1)(1+i_2)...(1+i_{20})$$

where i_k is the return in the k-th year.

You are given that for $k=1,2,...,20$:
(1) $i_1, i_2, ... i_{20}$ are independent;
(2) $(1+i_k)$ is lognormally distributed;
(3) The expected annual return $E(i_k)=5\%$; the standard deviation of the annual return $\sigma(i_k)=6\%$.

Find

(1) The expected value of investing \$1 after 20 years.
(2) The probability that the value of investing \$1 after 20 years is less than 95% of the expected value.

Solution

Many people have difficulty with this problem. The major difficulty is to determine whether $Y=(1+i_1)(1+i_2)...(1+i_{20})$ is normally distributed or lognormally distributed. To avoid the confusion, use Rule 1 and Rule 2.

We are told that $(1+i_k)$ is lognormal. This means that $\ln(1+i_k)$ is normal (Rule 2). Then $\ln Y=\ln(1+i_1)+\ln(1+i_2)+...+\ln(1+i_{20})$ is normal; the sum of several independent normal random variables is also normal.

If $\ln Y$ is normal, then $e^{\ln Y}=Y$ is lognormal (Rule 1). So Y is lognormal with the following parameters:

$$\ln Y=\ln(1+i_1)+\ln(1+i_2)+...+\ln(1+i_{20})$$

$$\Rightarrow \quad E[\ln Y]=20E[\ln(1+i)], \qquad Var[\ln Y]=20Var[\ln(1+i)]$$

The problem didn't give us $E\left[\ln(1+i)\right]$ and $Var\left[\ln(1+i)\right]$. However, it did give us $E(i)$ and $\sigma(i)$. So we need to find the $E\left[\ln(1+i)\right]$ and $Var\left[\ln(1+i)\right]$ using $E(i)$ and $\sigma(i)$.

$$E(X)=E(1+i)=1+E(i)=105\%$$
$$\sigma(X)=\sigma(1+i)=\sigma(i)=6\%$$

Solving the following equations:

$$\begin{cases} E(X)=e^{\mu_{\ln X}+\frac{1}{2}\sigma_{\ln X}{}^2}=105\% \\ Var(X)=\left(e^{\sigma_{\ln X}{}^2}-1\right)E^2(X)=(6\%)^2 \end{cases}$$

$$\Rightarrow \quad \sigma_{\ln X}=\sqrt{\ln\left[1+\frac{Var(X)}{E^2(X)}\right]}=\sqrt{\ln\left[1+\left(\frac{6\%}{105\%}\right)^2\right]}=5.7096\%$$

$$\mu_{\ln X}=\ln E(X)-\frac{1}{2}\sigma_{\ln X}{}^2=\ln 105\%-\frac{1}{2}(5.7096\%)^2=4.716\%$$

Then Y is lognormal with parameters

$$E[\ln Y]=20E[\ln X]=20(4.716\%)=0.9432$$
$$\sigma[\ln Y]=\sqrt{20}\sigma[\ln X]=\sqrt{20}(5.7096\%)=0.25534$$

$$E(Y)=e^{\mu_{\ln Y}+\frac{1}{2}\sigma_{\ln Y}{}^2}=e^{0.9432+\frac{1}{2}\left(0.25534^2\right)}=e^{0.9758}=2.6533$$

$$95\%E(Y)=(95\%)2.6533=2.5206$$

$$\Pr\left[Y<95\%E(Y)\right]=\Pr\left[Y<2.5206\right]=\Pr\left[\ln Y<\ln 2.5206\right]=\Phi\left[\frac{\ln 2.5206-E(\ln Y)}{\sigma(\ln Y)}\right]$$

$$=\Phi\left[\frac{\ln 2.5206-0.9432}{0.25534}\right]=\Phi(-0.0732)=1-\Phi(0.0732)=1-0.53=0.47$$

Homework for you: Rework all the problems in this chapter.

Chapter 26 Chi-square distribution

You have a χ^2 distribution when you square and sum n independent standard normal distributions.

Key points:

1. Let $Z_1, Z_2, ..., Z_n$ be independent standard normal distributions (i.e. each having mean of zero and standard deviation of one), then

   $X^2 = Z_1^2 + Z_2^2 + ... + Z_n^2$ where n is a positive integer.

   has a Chi-square $\chi^2{}_n$ distribution with n degrees of freedom.

   Please note that the random variable is X^2, not X.

2. Relationship between Chi-square $\chi^2{}_n$ distribution and gamma distribution (you can find the proof in some textbooks):

   Let $Y = X^2 = Z_1^2 + Z_2^2 + ... + Z_n^2$, where $Z_1, Z_2, ..., Z_n$ are independent standard normal distributions. Then Y has gamma distribution with parameters $\alpha = \dfrac{n}{2}$ and $\theta = 2$:

   $$f\left(Y = y\right) = \frac{1}{\theta^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}, \quad y \geq 0$$

3. Mean and variance formula:

   X^2 has gamma distribution with parameters $\alpha = \dfrac{n}{2}$ and $\theta = 2$. Then

   $E\left(X^2\right) = \alpha\theta = \left(\dfrac{n}{2}\right)2 = n$, (mean equal to the degree of freedom)

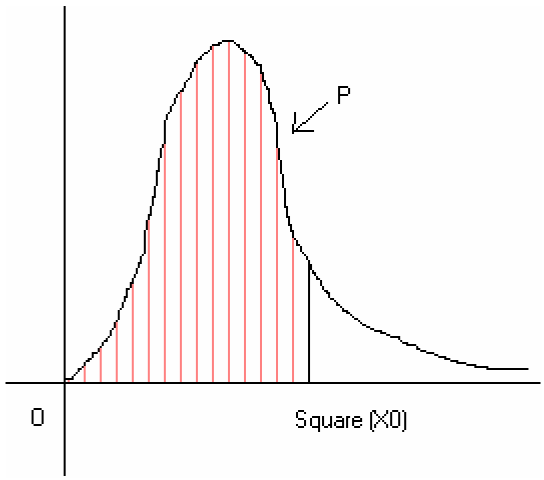
   $Var\left(X^2\right) = \alpha\theta^2 = \left(\dfrac{n}{2}\right)2^2 = 2n$

   (variance equal to twice the degree of freedom)

   $M\left(t\right) = \dfrac{1}{\left(1-2t\right)^{\frac{n}{2}}}$ (moment generating function)

4. $\chi^2{}_n$ is NOT symmetric (i.e. not bell-like). $\chi^2{}_n$ distribution is like a bell squashed from the right (i.e. the diagram is skewed to the left). As n increases, the mean and variance of $\chi^2{}_n$ increases and the $\chi^2{}_n$ diagram is less squashed on the right side (more bell-like).

5. You will need to know how to read the chi-square distribution table. A typical table looks like this:



P=the shaded area = $\Pr(x^2 < x_0{}^2)$

Sample numbers from a chi-square distribution table

			Value of P					
degree of freedom	0.005	0.010	0.025	0.050	0.950	0.975	0.990	0.995
1	0	0.00	0.00	0.00	3.84	5.02	0.64	7.88
2	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.60

Based on the above table, if the degree of freedom $n = 2$, then

$\Pr(x^2 < 5.99) = 0.95$, $\Pr(x^2 < 7.38) = 0.975$

Sample Problems and Solutions

Problem 1

A chi-square random variable Y^2 has a variance of 10.

You are given the following table:

			Value of P					
degree of freedom	0.005	0.010	0.025	0.050	0.950	0.975	0.990	0.995

5	0.41	0.55	0.83	1.15	11.07	12.83	15.09	16.75
10	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19

Find the mean;
Find the 5th percentile of Y^2;
Given that $\Pr(a \le Y^2 \le b) = 0.005$, find $b - a$.

Solution

The variance of a chi-square distribution is equal to twice the degree of freedom n.

$$2n = 10 \quad \Rightarrow n = 5$$

The mean of a chi-square distribution is equal to the degree of freedom. So the mean is 5.

To find the 5th percentile, we need to solve the equation

$\Pr(Y^2 \le \text{5th percentile}) = 5\%$.

From the given Chi-square distribution table, we find that value that corresponds to $n = 5$ and $p = 0.05$ is 1.15. Thus, the 5th percentile is 1.15.

There are many combinations of a, b that satisfy $\Pr(a \le Y^2 \le b) = 0.005$. However, based on the given chi-square distribution, only two combinations satisfy $\Pr(a \le Y^2 \le b) = 0.005$ for $n = 5$:

$$(a, b) = (0,\ 0.41) \text{ or } (a, b) = (0.41,\ 0.55)$$

So we have

$$b - a = 0.41 \text{ or } b - a = 0.14$$

Please note that the information for $n = 10$ in the chi-square distribution table is not needed for this problem.

Problem 2

A chi-square random variable Y^2 has a mean of 1.
No chi-square distribution table is given.

Find $\Pr(Y^2 > 1.21)$

Solution

The mean of a chi-square distribution is equal to the degree of freedom. So the degree of freedom is one.

If a chi-square distribution Y^2 has a freedom of one, then Y is simply a standard normal random variable (based on the definition of chi-square distribution).

$$\Pr(Y^2 > 1.21) = 1 - \Pr(Y^2 \le 1.21) = 1 - \Pr(-\sqrt{1.21} \le Y \le \sqrt{1.21})$$
$$\Pr(-\sqrt{1.21} \le Y \le \sqrt{1.21}) = \Pr(-1.1 \le Y \le 1.1) = \Phi(1.1) - \Phi(-1.1) = \Phi(1.1) - [1 - \Phi(1.1)]$$
$$\Pr(-\sqrt{1.21} \le Y \le \sqrt{1.21}) = 2\Phi(1.1) - 1 = 2(0.8643) - 1 = 0.7286$$

$$\Pr(Y^2 > 1.21) = 1 - \Pr(-\sqrt{1.21} \le Y \le \sqrt{1.21}) = 1 - 0.7286 = 0.2714$$

Problem 3

Y has Chi-square distribution with $n = 6$. Find $\Pr(Y \le 10.15)$.

Solution

Y has gamma distribution with parameters $\alpha = \frac{n}{2} = 3$ and $\theta = 2$.

$\Pr(Y \le 10.15) = \Pr(\text{it takes time } 10.15 \text{ or less to have 3 random events})$
=Pr(the # of events that occurred during $[0,\ 10.15] = 3, 4, 5, \ldots + \infty$)
= 1- Pr(the # of events that occurred during $[0,\ 10.15] = 0, 1, 2$)

$$= 1 - \underbrace{e^{-\frac{10.15}{2}}\left[1 + \frac{\frac{10.15}{2}}{1!} + \frac{\left(\frac{10.15}{2}\right)^2}{2!}\right]}_{\text{Poisson distribution}} = 88.15\%$$

Please note that if the degree of freedom is an odd number such as $n = 7$, then Y has gamma distribution with parameters $\alpha = \frac{7}{2} = 3.5$ and $\theta = 2$.

Because α is not an integer, we cannot do the following (the number of events must be an integer for Poisson distribution to work):

$\Pr\left(Y \leq 10.15\right) = \Pr$(it takes time 10.15 or less to have 3.5 random events)
=Pr(the # of events that occurred during $[0,10.15] = 3.5,\ 4.5,\ ...+\infty$)

If $n = 7$, then we have to calculate $\Pr\left(Y \leq 10.15\right)$ through other means such as using a gamma distribution table.

Homework for you: Rework all the problems in this chapter.

Chapter 27 Bivariate normal distribution

If two normal random variables $X \sim N(E(X), \sigma_X)$ and $Y \sim N(E(Y), \sigma_Y)$ have a correlation coefficient ρ $(-1 \le \rho \le 1)$, then the joint distribution of X, Y is a bivariate normal distribution with the following pdf:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times$$

$$\exp\left\{ -\frac{1}{2\left(1-\rho^2\right)} \left[\left(\frac{x-E(X)}{\sigma_X} \right)^2 - 2\rho \frac{\left(x-E(X)\right)\left(y-E(Y)\right)}{\sigma_X\sigma_Y} + \left(\frac{y-E(Y)}{\sigma_Y} \right)^2 \right] \right\}$$

where $-\infty \le X \le +\infty, \quad -\infty \le Y \le +\infty$

No need to memorize this complex formula.

The derivation of this pdf is very complex and involves the transformation of a joint distribution function. I recommend that you do not bother learning how to derive the above formula.

If we integrate over Y, we get the X-*marginal distribution*:

$$f(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x-E(X)}{\sigma_X} \right)^2 \right)$$

You should recognize this pdf. It is just a regular pdf of a normal random variable with the mean $E(X)$ and the standard deviation σ_X.

If we integrate over X, we get the Y-*marginal distribution*:

$$f(y) = \int_{-\infty}^{\infty} f(x,y)dx = \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y-E(Y)}{\sigma_Y} \right)^2 \right)$$

This is just a regular pdf of a normal random variable with a mean $E(Y)$ and a standard deviation σ_Y.

The conditional distribution of Y given X = x also has a normal distribution:

$$f_{Y|X=x}(Y|X=x)=\frac{1}{\sigma_{Y|X=x}\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{y-E(Y|X=x)}{\sigma_{Y|X=x}}\right)^2\right]$$

Where

$$E(Y|X=x)=E(Y)+\rho\frac{\sigma_Y}{\sigma_X}\left[x-E(X)\right]=E(Y)+\frac{Cov(X,Y)}{\sigma_X\sigma_Y}\frac{\sigma_Y}{\sigma_X}\left[x-E(X)\right]$$

$$=E(Y)+\frac{Cov(X,Y)}{Var(X)}\left[x-E(X)\right]$$

$$\sigma^2_{Y|X=x}=(1-\rho^2)\sigma^2_Y.$$

You will want to memorize the conditional mean and variance formulas. It would be easy for an exam question to be written on these.

If we examine the formula $E(Y|X=x)=E(Y)+\rho\frac{\sigma_Y}{\sigma_X}\left[x-E(X)\right]$, we realize that $E(Y|X=x)$ and x have a simple linear relationship. In other words, if X is fixed at x, the conditional mean of Y is simply a linear function of x:

$$E(Y|X=x)=ax+b,\ \text{where}\ a=\rho\frac{\sigma_Y}{\sigma_X},\ \mathrm{b}=\mathrm{E}(\mathrm{Y})-\rho\frac{\sigma_Y}{\sigma_X}\mathrm{E}(\mathrm{X})$$

The conditional distribution of X given Y = y also has a normal distribution:

$$f_{X|Y=y}(X|Y=y)=\frac{1}{\sigma_{X|Y=y}\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{y-E(X|Y=y)}{\sigma_{X|Y=y}}\right)^2\right]$$

Where

$$E(X|Y=y)=E(X)+\rho\frac{\sigma_X}{\sigma_Y}\left[y-E(Y)\right]=E(X)+\frac{Cov(X,Y)}{Var(Y)}\left[y-E(Y)\right]$$

$$\sigma^2_{Y|\,X=x}=(1-\rho^2)\sigma^2_Y.$$

Similarly, $E(X|Y=y)$ and y have a simple linear relationship.

Please note the conditional mean and variance formulas are symmetric in terms of X, Y. If you have memorized the conditional mean and variance formulas for $Y|X=x$, you can get the conditional mean and variance formulas for $X|Y=y$ by switching X, Y.

For a bivariate normal distribution, if $\rho = 0$, the pdf becomes

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y}\exp\left\{-\frac{1}{2}\left[\left(\frac{x-E(X)}{\sigma_X}\right)^2 + \left(\frac{y-E(Y)}{\sigma_Y}\right)^2\right]\right\}$$
$$= \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left[-\frac{1}{2}\left(\frac{x-E(X)}{\sigma_X}\right)^2\right] \times \frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left[-\frac{1}{2}\left(\frac{y-E(Y)}{\sigma_Y}\right)^2\right] = f(x)f(y)$$

Then X, Y are independent. So $\rho = 0 \quad \Leftrightarrow \quad X, Y$ are independent.

For a non-bivariate normal distribution, typically $\rho = 0$ does not guarantee that X, Y are independent. However, if X, Y are normal random variables, then $\rho = 0 \quad \Rightarrow \quad X, Y$ are independent. This is an exception to the general rule that if $\rho = 0$ then X, Y are not guaranteed to be independent.

However, for any distribution (whether bivariate or not), if X, Y are independent, then $\rho = 0$.

Sample Problems and Solutions

Problem One

Let Z_1, Z_2 represent two independent standard normal random variables. Let $X = 2Z_1 + 3Z_2 + 4$, $Y = 2Z_1 - 5Z_2 + 2$.

Find $E(Y|X=0), Var(Y|X=0)$.

Solution

You should have memorized the following formulas:

$$E(Y|X=x)=E(Y)+\frac{Cov(X,Y)}{Var(X)}\left[x-E(X)\right]$$
$$\sigma^2_{Y|X=x}=(1-\rho^2)\sigma^2_Y.$$

Otherwise, the solution to this problem becomes pure guesswork.

The first step is to calculate ρ.

$$\rho=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

We'll use the following general formulas:

$$\boxed{\begin{aligned}&Cov(aX+bY+e,cX+dY+f)=acVar(X)+(ad+bc)Cov(X,Y)+bdVar(Y)\\&Var(aX+bY)=Var(aX)+Var(bY)+2abCov(X,Y)\\&Var(aX)=a^2\,Var(X)\end{aligned}}$$

$$Cov(X,Y)=Cov(2Z_1+3Z_2+4,\,2Z_1-5Z_2+2)=Cov(2Z_1+3Z_2,\,2Z_1-5Z_2)$$

The above evaluation stands because the covariance between a constant and any random variable is always zero.

$$\begin{aligned}Cov(X,Y)&=Cov(2Z_1+3Z_2,\,2Z_1-5Z_2)\\&=Cov(2Z_1,2Z_1)+Cov(2Z_1,-5Z_2)+Cov(3Z_2,2Z_1)+Cov(3Z_2,-5Z_2)\end{aligned}$$

$$Cov(2Z_1,2Z_1)=4Var(Z_1)=4$$
$$Cov(3Z_2,-5Z_2)=-15Var(Z_2)=-15$$
$$Cov(2Z_1,-5Z_2)=-10Cov(Z_1,Z_2)=0 \quad (Z_1,Z_2 \text{ are independent})$$
$$Cov(3Z_2,2Z_1)=6Cov(Z_1,Z_2)=0$$

$$Cov(X,Y)=4-15=-11$$

$$Var(X)=Var(2Z_1+3Z_2+4)=4Var(Z_1)+9Var(Z_2)=4+9=13,\ \sigma_X=\sqrt{13}$$
$$Var(Y)=Var(2Z_1-5Z_2+2)=4Var(Z_1)+25Var(Z_2)=4+25=29,\ \sigma_Y=\sqrt{29}$$

$$\rho=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}=-\frac{11}{\sqrt{13}\times\sqrt{29}}$$

$$E(X)=E(2Z_1+3Z_2+4)=4,\ E(Y)=E(2Z_1-5Z_2+2)=2$$

Next, simply apply the memorized formulas:

$$E(Y|X=x)=E(Y)+\frac{Cov(X,Y)}{Var(X)}\left[x-E(X)\right],\quad \sigma^2_{Y|X=x}=(1-\rho^2)\sigma^2_Y.$$

$$E\left(Y|X=0\right)=2+\frac{-11}{13}\left(0-4\right)=\frac{70}{13}$$

$$\sigma^2_{Y|X=0}=\left[1-\left(-\frac{11}{\sqrt{13}\times\sqrt{29}}\right)^2\right]\times 29=\frac{256}{13}$$

Homework for you: Rework the problem in this chapter.

Chapter 28 Joint density and double integration

Joint density problems are among the more difficult problems commonly tested in Exam P. Many candidates dread these types of problems. To score a point, not only do you need to know probability theories, but you must also be precise and quick at doing double integration.

The good news is that there is a generic approach to this type of problem. Once you understand this generic approach and do some practice problems, double integration and joint density problems are just another source of routine problems where you can easily score points in the exam.

Before finding a general approach to joint density problems, let's go back to the basic idea behind double integration.

Basics of integration

Problem 1 (discrete random variable)

A random variable X has the following distribution:

X	f (X = x)
0	.25
1	.25
2	.25
3	.25

Find $E(X)$.

Solution

maximum of x is 3

$$E(X)=\sum_{x=0}^{3} xf(x)=0(.25)+1(.25)+2(.25)+3(.25)=1.5 \qquad \text{(Equation 1)}$$

minimum of x is 0

Problem 2 (Continuous random variable)

A random variable X has the following probability distribution.

$$f(x) = \frac{2}{9}x, \quad 0 \le x \le 3$$

Find $E(X)$.

Solution

maximum of x is 3

$$E(X) = \int_0^3 xf(x)dx = \int_0^3 x(\frac{2}{9}x)\,dx = \frac{2}{9}(\frac{1}{3}x^3)\Big|_0^3 = 2 \qquad \text{(Equation 2)}$$

minimum of x is 0

You see that Equation 1 and Equation 2 are very similar. In both places, we add up $xf(x)$ over all possible values of x ranging from the minimum to the maximum. The only difference is that we use summation for Equation 1 (because we have a finite number of x's) and integration for Equation 2 (because we have an infinite number of x's).

What about finding a mean for a joint distribution? Two random variables X, Y have a joint distribution $f(x, y)$? Can you guess the formula for $E(X)$?

The formula is:

$$E(X) = \underbrace{\int_{\min y}^{\max y} \overbrace{\int_{\min x}^{\max x} xf(x,y) \quad dx}^{X} \quad dy}_{Y} \qquad \text{(Equation 3)}$$

You see that Equation (3) is very similar to Equation (1) and (2). The only difference is that now we have double integration because we have to sum things up twice. First, we integrate $xf(x,y)$ over all possible values of X (inner integration) by holding y constant. Then, we integrate over all possible values of Y.

You can also write

$$E(X) = \underbrace{\int_{\min x}^{\max x} \overbrace{\int_{\min y}^{\max y} xf(x,y) \quad dy}^{Y} \quad dx}_{X} \quad \text{(Equation 4)}$$

This time, the inner integration sums over y from min of y to max of y, and the outer integration sums over x from min of x to max of x.

You should never write (conceptually wrong):

$$E(X) = \underbrace{\int_{\min x}^{\max x} \overbrace{\int_{\min y}^{\max y} xf(x,y) \qquad dx}^{\min/\max\ Y \text{ and } dx \text{ don't match}} \quad dy}_{\min/\max\ X \text{ and } dy \text{ don't match}}$$

In other words, if your inner integration is dx, then the inner integration must sum over x from min x to max x; the outer integration must sum over y from min y to max y.

Nor should you write:

$$E(X) = \underbrace{\int_{\min y}^{\max y} \overbrace{\int_{\min x}^{\max x} xf(x,y) \qquad dy}^{\min/\max\ X \text{ and } dy \text{ don't match}} \quad dx}_{\min/\max\ Y \text{ and } dx \text{ don't match}}$$

Now we are ready to tackle the joint density problems and related double integrations.

General approach to joint density problems

To illustrate the general approach, let's use a simple example:

Problem 3

The total car-related damage (measured in thousands of dollars) incurred by an auto insurance policyholder in an auto accident can be classified into two categories: X, which is the damage to his own car; and Y, which is the damage to the other driver's car. X, Y have a joint density of $f(x, y)$. What is the probability that the total loss does not exceed 2 (thousand dollars)?

Step One –Draw a 2-D region for all the possible combinations of two random variables (X,Y). You need this 2-D region for integration.

To begin with, your 2-D region should be where the joint density function $f(x,y)$ exits; any data points (x,y) where $f(x,y)$ is undefined will be outside the 2-D region. Any additional constraint on (X,Y) will shrink the 2-D region.

In this problem, obviously $f(x,y)$ exists only in $x \geq 0, y \geq 0$ (i.e. the first quadrant). So your 2-D region is now the first quadrant before any additional constraints (the shaded area in Figure 1).

The additional constraint is $x + y \leq 2$ (total loss not exceeding 2). Where is $x + y \leq 2$ or $y \leq 2 - x$? You should remember this rule:

$y \geq f(x)$ lies above $y = f(x)$ because y is equal to or greater than $f(x)$; $y \leq f(x)$ lies below $y = f(x)$ because y is equal to or less than $f(x)$.

$x + y \leq 2$ is the shaded area in Figure 2.

Please also note that in Figure 2, the shaded area is also for $x + y < 2$. In other words, there's no difference between $x + y \leq 2$ and $x + y < 2$. Why? Double integration integrates over an area; a single point or a line doesn't have any area.

Put another way, the joint density function $f(x,y)$ is zero on a single point or a line. $f(x,y)$ is meaningful only if you integrate $f(x,y)$ over an area.

This is similar to the concept that $f(x)$, the density function of a continuous univariate variable X, is zero at any single point. $f(x)$ is meaningful only if you integrate $f(x)$ over a line -- for example, if you are using $f(x)$ to get the accumulative density function of $F(x)$.

In general, for joint density problems, the 2-D region for $y \geq f(x)$ is identical to the 2-D region for $y > f(x)$; the 2-D region for $y \leq f(x)$ is identical to the 2-D region for $y < f(x)$;

Finally, the 2D region for $x \geq 0, y \geq 0$ and $x + y \leq 2$ should be the intersection of the shaded region for $x \geq 0, y \geq 0$ in Figure 1 and the shaded region for $x + y \leq 2$ in Figure 2. This is the shaded triangle AOB in Figure 3.

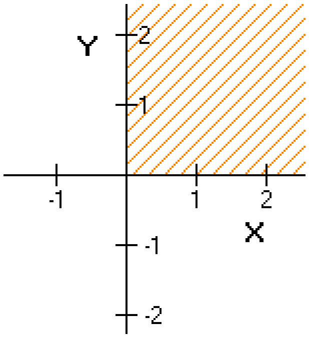


Figure 1
Shaded area= $x \geq 0, y \geq 0$

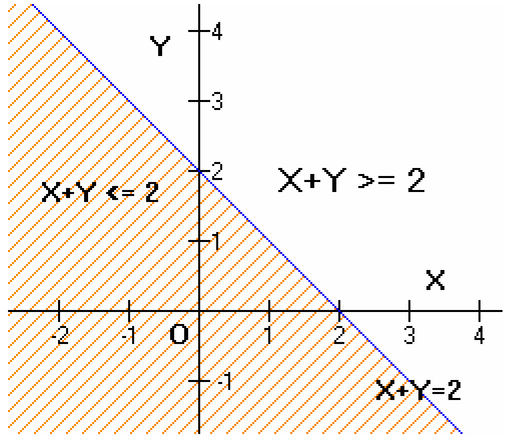


Figure 2
Shaded area= $x + y \leq 2$

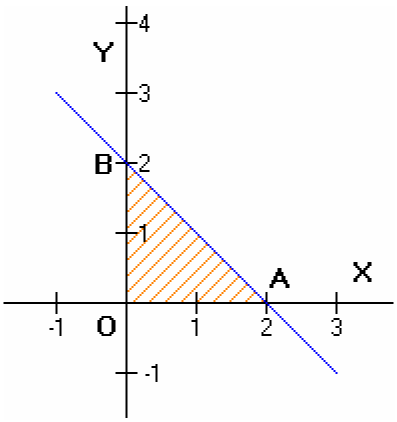


Figure 3 Shaded area = $(x \geq 0, y \geq 0) \cap (x + y \leq 2)$

Step Two --- Choose one variable for the outer integration. Always set up the outer integration first.

You can choose either X or Y for outer integration.

If you choose X for outer integration, you can set up the following double integration:

$$\Pr(x+y \leq 2) = \underbrace{\int_{\substack{\text{MIN of } X \\ \text{in the 2-D} \\ \text{region}}}^{\substack{\text{MAX of } X \\ \text{in the 2-D} \\ \text{region}}} \left(\int f(x,y)\, dy \right) dx}_{\substack{\text{Outer Integration} \\ \text{Always set up outer integration first}}}$$

You can clearly see that in the 2-D region $(x \geq 0, y \geq 0) \cap (x + y \leq 2)$, the minimum of X =0 (at O, the center of the plane) and maximum of X =2 (at point A). So we have:

$$\Pr(x+y\leq 2)=\underbrace{\int_0^2 \left(\int f(x,y)\,dy\right) dx}_{\substack{\text{Outer Integration}\\ \text{Always set up outer integration first}}}$$

Of course, you can choose Y for outer integration:

$$\Pr(x+y\leq 2)=\underbrace{\int_{\substack{\text{MIN of } Y\\ \text{in the 2-D}\\ \text{region}}}^{\substack{\text{MAX of } Y\\ \text{in the 2-D}\\ \text{region}}} \left(\int f(x,y)\,dx\right) dy}_{\substack{\text{Outer Integration}\\ \text{Always set up outer integration first}}}$$

In the 2-D region $(x \geq 0, y \geq 0) \cap (x + y \leq 2)$, the minimum of Y =0 (at O, the center of the plane) and maximum of Y =2 (at point B). So we have:

$$\Pr(x+y\leq 2)=\underbrace{\int_0^2 \left(\int f(x,y)\,dx\right) dy}_{\substack{\text{Outer Integration}\\ \text{Always set up outer integration first}}}$$

Step 3—find the min/max of the inner integration (a little tricky).

Here is a general rule:

If the inner integration is on Y, we draw a vertical line $X = x$ that cuts through the 2-D region. The Y coordinates of the two cutting points are the lower bound and upper bound for the inner Y-integration. If the inner integration is on X, we draw a horizontal line $Y = y$ that cuts through the 2-D region. The X coordinates of the two cutting points are the lower bound and upper bound for the inner X-integration.

Let's apply the above rule. For inner Y-integration, we draw a vertical line $X = x$ that cuts through the 2-D region. The two cutting points are $C(x, 0)$ and $D(x, 2 - x)$. So we integrate Y over the line CD using C (whose Y coordinate is zero) as the lower bound and D (whose Y coordinate is $2 - x$) as the upper bound.

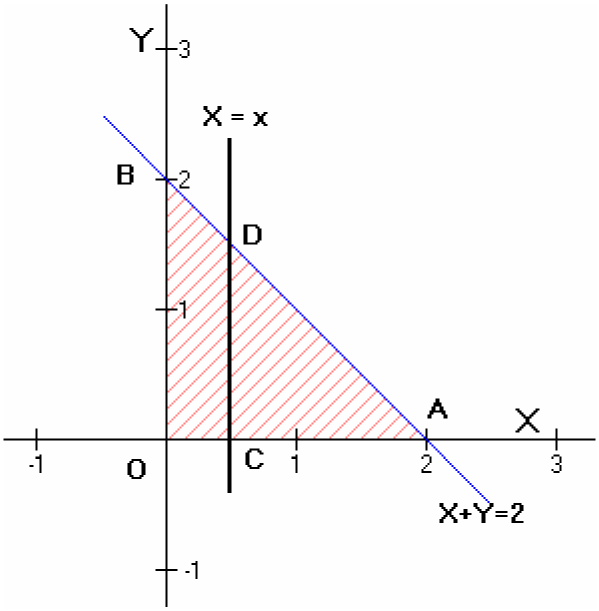


Figure 4

$$\Pr(x+y\le 2)=\int_0^2\left(\underbrace{\int f(x,y)\,dy}_{\text{Integrate over CD}}\right)dx=\int_0^2\left(\int_0^{2-x} f(x,y)\,dy\right)dx$$

Line CD represents all the possible values of Y in the 2-D region given that $X=x$. So the line CD is really $Y\,|\,X=x$. We will revisit this point later on when we need to calculate $f(x)$, the X-marginal density of $f(x,y)$.

So in the above equation, the inner integration is done over all possible values of Y given $X=x$. Then, in the outer integration, we integrate over all possible values of X. The inner and outer integration work together, scanning everywhere within the 2-D region for (X,Y), and summing up the total probability within the 2-D region.

Similarly, for the inner X-integration, we draw a horizontal line $Y=y$ that cuts through the 2-D region. The two points are $E(0,y)$ and $F(2-y,y)$. So we integrate X over the line EF, using E (whose X coordinate is zero) as the lower bound and F (whose X coordinate is $2-y$) as the upper bound.

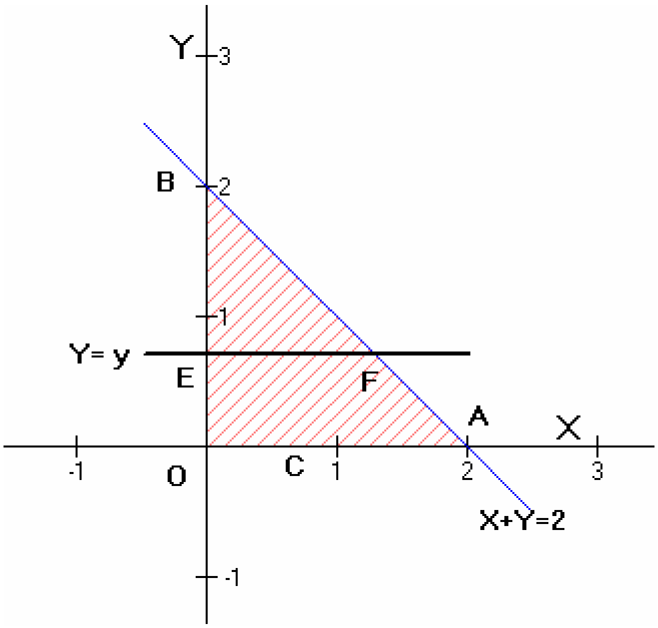


Figure 5

$$\Pr(x+y\le 2)=\int_0^2\left(\underbrace{\int f(x,y)\,dx}_{\text{Integrate over line EF}}\right)dy=\int_0^2\left(\int_0^{2-y} f(x,y)\,dx\right)dy$$

Line EF represents all the possible values of X in the 2-D region given that $Y=y$. So the line CD is really $X|Y=y$. We will also revisit this point later on when we need to calculate $f(y)$, the Y-marginal density of $f(x,y)$.

So in the above equation, the inner integration is done over all possible values of X given $Y=y$. Then the outer integration is done over all possible values of Y. The inner and outer integration work together, scanning everywhere within the 2-D region for (X,Y), and summing up the total probability within the 2-D region.

Now we have set up a complete double integration:

$$\Pr(x+y\le 2)=\int_0^2\left(\int_0^{2-x} f(x,y)\,dy\right)dx$$

Or $\quad \Pr(x+y\le 2)=\int_0^2\left(\int_0^{2-y} f(x,y)\,dx\right)dy$

Step 4 (final step) – evaluate the double integration starting from the inner integration.

$$\Pr(x+y\le 2)=\underbrace{\int_0^2 \left(\underbrace{\int_0^{2-x} f(x,y)\,dy}_{\text{Do inner first}}\right) dx}_{\text{Do outer last}}$$

Or $$\Pr(x+y\le 2)=\underbrace{\int_0^2 \left(\underbrace{\int_0^{2-y} f(x,y)\,dx}_{\text{Do inner first}}\right) dy}_{\text{Do outer last}}$$

Let's summarize the 4 steps for tackling joint density and double integration problems:

Step 1 – Draw a 2-D region. Initially, the 2-D region is where the joint density $f(x,y)$ exists. Any additional constraint shrinks the 2-D region.
When drawing the 2-D region, remember that $y \ge f(x)$ is above $y = f(x)$ and $y \le f(x)$ is below $y = f(x)$.

Step 2 – Set up the outer integration first. Choose either X or Y as the inner integration variable. Find the min/max values of the outer integration variable in the 2-D region. They are the lower and upper bounds of the outer integration.

Step 3 – Set up the inner integration. To determine the lower and upper bounds, draw a vertical line (if the inner integration is on Y) or horizontal line (if the inner integration is on X) that cuts through the 2-D region. Use the two cutting points as the lower and upper bounds for the inner integration.

Step 4 – Evaluate the double integration. Do the inner integration first.

Let's practice the above steps.

Problem 4

Random variables X and Y have the following joint distribution:

$$f(x,y)=\begin{cases} k\left(1-x+2y\right), & \text{for } 0\le x\le y\le 2-x \\ 0, & \text{elsewhere} \end{cases}$$

Find $E(X)$.

Solution

First, we need to solve k by using the equation:

$$\iint_{0 \le x \le y \le 2-x} f(x,y)\,dx\,dy = 1$$

Step One --- Determine the 2-D region

We break down $0 \le x \le y \le 2-x$ into three conditions:

$x \ge 0$ (1st and 4th quadrants.)

$x \le y$ (Figure 6)

$y \le 2-x$ (Figure 7)

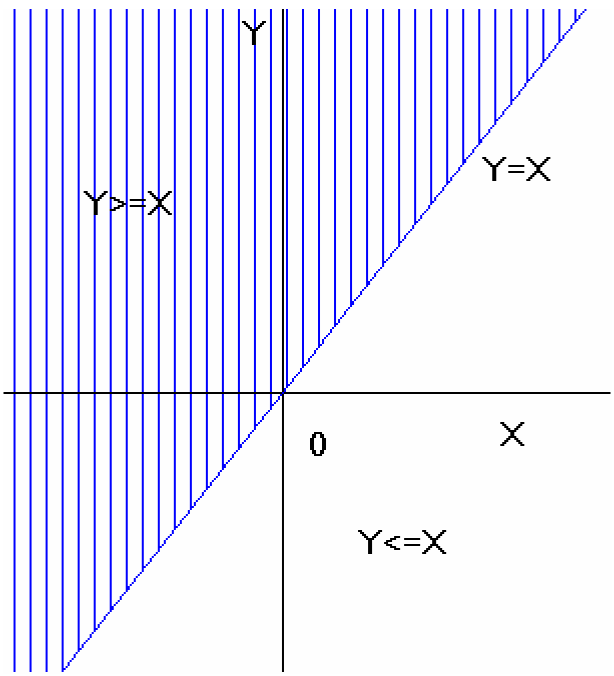


Figure 6 Shaded area=Y $\ge$ X

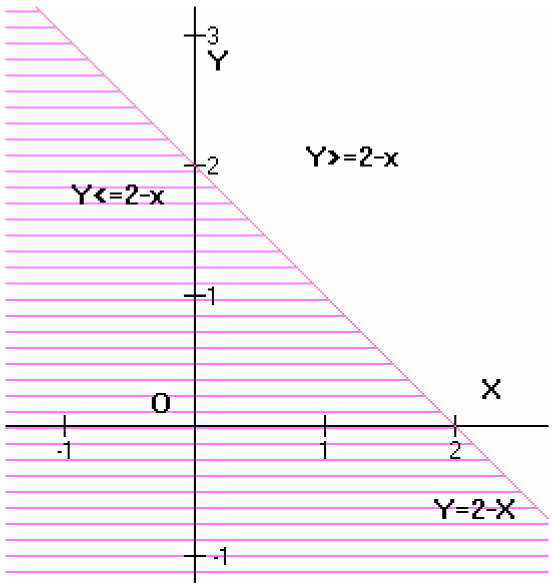


Figure 7 Shaded area=Y $\leq 2-X$

Now if you put all the separate constraints together, you'll find the desired 2-D region for $0 \leq x \leq y \leq 2-x$ (see the shaded area in Figure 8):

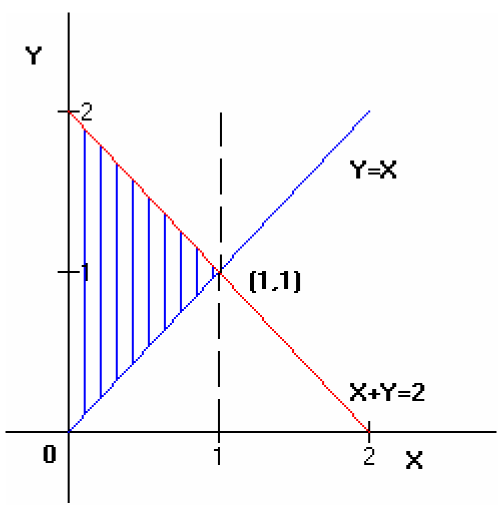


Figure 8 shaded area= $0 \leq x \leq y \leq 2-x$

Step Two ---Set up the outer integration

If we do the outer integration on X :

In the 2-D region, min X =0, max X =1.

$$\iint_{0 \le x \le y \le 2-x} f(x,y)\,dx\,dy = \int_0^1 \left(\int f(x,y)dy \right) dx$$

If we do the outer integration on Y :

In the 2-D region, min Y =0, max Y =2.

$$\iint_{0 \le x \le y \le 2-x} f(x,y)\,dx\,dy = \int_0^2 \left(\int f(x,y)dx \right) dy$$

Step Three --- Determine the lower and upper bounds for the inner integration.

If we do the inner integration on Y :

To find the lower and upper bounds on Y , we draw a vertical line $X = x$ that cuts through the 2-D region. The two cutting points are M($x, 2-x$) and N(x, x). So the lower bound is x (the Y coordinate of N) and the upper bound is $2-x$ (the Y coordinate of M). See Figure 9.

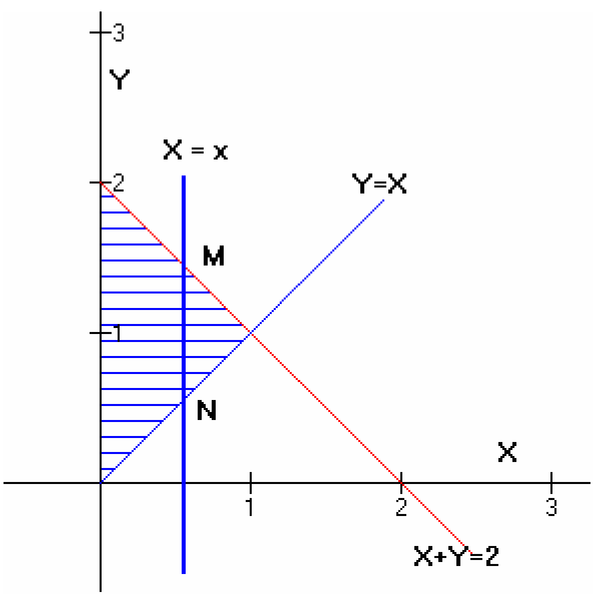


Figure 9

$$\int_0^1 \left(\int f(x,y)dy \right) dx = \underbrace{\int_0^1 \overbrace{\int_{NM} f(x,y)\ dy}^{Y}\ dx}_{X} = \underbrace{\int_0^1 \overbrace{\int_x^{2-x} f(x,y)\ dy}^{Y}\ dx}_{X}$$

If we do inner integration on X (more difficult):

(See Figure 10) To determine the lower and upper bounds of X, we need to draw two horizontal lines DE and GH that cut through the 2-D region. This is because the upper bound of X for the triangle ABC (the top portion of the 2-D region) is different from the upper bound of X for the triangle ABO (the lower portion of the 2-D region).

For the line D$(0, y)$ and E$(2-y, y)$, the lower bound of X is zero (the X coordinate of D) and the upper bound is $2-y$ (the X coordinate of E); for the line G$(0, y)$ and H(y, y), the lower bound is zero (the X coordinate of G) and the upper bound is y (the X coordinate of H).

As a result, to do the inner and outer integrations, we need to divide the 2-D region into two sub-regions: ABO and ABC.

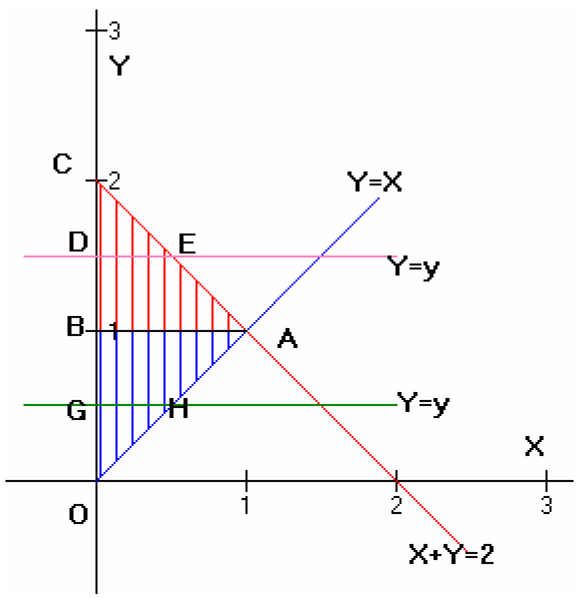


Figure 10

So the double integration becomes:

$$\int_0^2 \left(\int f(x,y)dx \right) dy = \underbrace{\int_0^1 \int_{GH} f(x,y)dx\,dy}_{\triangle ABO} + \underbrace{\int_1^2 \int_{DE} f(x,y)dx\,dy}_{\triangle ABC}$$

$$= \underbrace{\int_0^1 \int_0^y f(x,y)dx\,dy}_{\triangle ABO} + \underbrace{\int_1^2 \int_0^{2-y} f(x,y)dx\,dy}_{\triangle ABC}$$

Step Four – Evaluate the integration starting from the inner integration.

We'll do the outer integration on X and the inner integration on Y (this is easier).

$$\underbrace{\int_0^1 \overbrace{\int_x^{2-x} f(x,y)\ dy}^{Y}\ dx}_{X} = 1 \quad \Rightarrow \int_0^1 \int_x^{2-x} k(1-x+2y)dydx = 1$$

We always do the inner integration first.

$$\int_x^{2-x} k(1-x+2y)dy \quad \text{(remember that } x \text{ is a constant here)}$$

To help us remember that we are integrating over y, not x, we underline any item that has y in it, treating all other items as constants (not underlined):

$$\Rightarrow \int_x^{2-x} k(1-x+2\underline{\underline{y}})\underline{\underline{dy}} = k(\underline{\underline{y}} - x\underline{\underline{y}} + \underline{\underline{y^2}})\Big|_x^{2-x}$$

$$= k\left\{ \left[(2-x) - x(2-x) + (2-x)^2\right] - \left[x - x^2 + x^2\right] \right\} = k(2x^2 - 8x + 6)$$

Next, we integrate $k(2x^2 - 8x + 6)$ over x:

$$\int_0^1 \int_x^{2-x} k(1-x+2y)dydx = \int_0^1 k(2x^2 - 8x + 6)dx$$

$$= k\left(\frac{2}{3}x^3 - 4x^2 + 6x\right)\Big|_0^1 = k\left(\frac{2}{3} - 4 + 6\right) = \frac{8}{3}k$$

$$\Rightarrow \quad \frac{8}{3}k = 1 \quad \Rightarrow k = \frac{3}{8}$$

Now we are ready to find $E(X)$.

$$E(X)=\int_0^1\int_x^{2-x} x\frac{3}{8}(1-x+2y)dydx$$

First do the inner integration:

$$\int_x^{2-x} x\frac{3}{8}(1-x+2y)dy=\frac{3}{8}x(2x^2-8x+6)=\frac{3}{8}(2x^3-8x^2+6x)$$

Next do the outer integration:

$$E(X)=\int_0^1\frac{3}{8}(2x^3-8x^2+6x)dx=\frac{3}{8}\left(\frac{2}{4}x^4-\frac{8}{3}x^3+\frac{6}{2}x^2\right)\Bigg|_0^1$$

$$=\frac{3}{8}\left(\frac{2}{4}-\frac{8}{3}+\frac{6}{2}\right)=\frac{5}{16}$$

Here's a repeat of an exam hint I said before. In the heat of the exam, converting $\frac{3}{8}(\frac{2}{4}-\frac{8}{3}+\frac{6}{2})$ into a neat fraction is painful and prone to errors. You should use the calculator technique I showed you in Chapter 2 and quickly and accurately use your calculator to deal with fractions.

Problem 5 (continue Problem 4)

Random variables X and Y have the following joint distribution:

$$f(x,y)=\begin{cases}\frac{3}{8}(1-x+2y), & \text{for } 0\le x\le y\le 2-x\\ 0, & \text{elsewhere}\end{cases}$$

Find $Var(X)$.

Solution

If you can accurately find the 2-D region and correctly set up the double integration form, statistic formulas on joint distributions are similar to the formulas for single integration ---except we have to use a joint pdf (instead of a single pdf) and integrate over a 2-D region.

$Var(X)=E(X^2)-E^2(X)$ (same formula)

We already know that $E(X)=\dfrac{5}{16}$

$$E(X^2)=\int\int x^2 f(x,y)dxdy \qquad \text{(similar formula)}$$

$$=\int_0^1\int_x^{2-x} x^2\frac{3}{8}(1-x+2y)dy\,dx=\frac{3}{8}\int_0^1 x^2(2x^2-8x+6)dx=\frac{3}{20}$$

$$Var(X)=\frac{3}{20}-\left(\frac{5}{16}\right)^2\approx 0.05234$$

If the problem asks us to find other statistics such as $Cov(X,Y)$, then

$$E(XY)=\int\int xyf(x,y)dxdy=\int_0^1\int_x^{2-x} xy\frac{3}{8}(1-x+2y)dy\,dx=....$$

$$E(Y)=\int\int yf(x,y)dxdy=\int_0^1\int_x^{2-x} y\frac{3}{8}(1-x+2y)dy\,dx=...$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

What about word problems about a joint pdf? Most of the difficulty in such word problems is about finding the right 2-D region. If you can correctly identify the 2-D region, the rest of the work is just (tedious) integration. Let's have some word problems and practice how to pinpoint the 2-D region for double integration.

Problem 6 (word problem)

Claims on homeowners insurance consist of two parts – claims on the main dwelling (i.e. house) and claims on other structures on the property (such as a garage). Let X be the portion of a claim on the house, let Y be the portion of the same claim on the other structures.

X,Y have the joint density function

$$\begin{cases} f(x,y) & \text{if } 0<x<2,\quad 0<y<3 \\ 0 & \text{if elsewhere} \end{cases}$$

Find the probability that the house portion of the claim is more than two times of the other structures portion of the same claim.

Solution

"The probability that the house portion of the claim is more than two times of the other structures portion of the same claim" is $\Pr(X > 2Y)$.

I will only show you how to find the 2-D region and how to set up the double integration. You can do the actual integration. Don't worry about how the joint density function, $f(x,y)$, looks because $f(x,y)$ doesn't affect the 2-D region.

$0 < x < 2, \quad 0 < y < 3$ is an obvious constraint. The other constraint is that the claim on the house must be more than two times of the claim on the other structures (i.e. $X > 2Y$ or $Y < 0.5X$). So the 2-D region is formed by:

$0 < x < 2, \quad 0 < y < 3, Y < 0.5X$

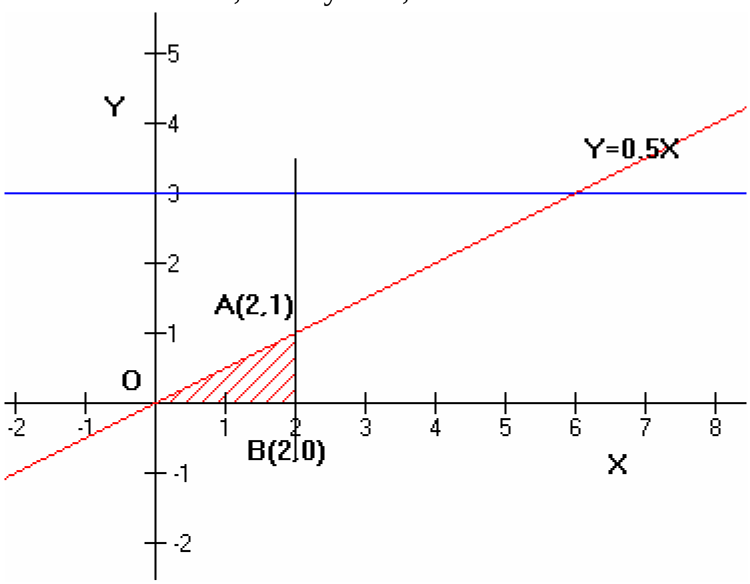


Figure 11

The shaded triangle AOB in Figure 11 is the 2-D region for integration.

$$\Pr(X > 2Y) = \int_0^2 \int_0^{0.5x} f(x,y)\,dy\,dx$$

Problem 7

Claims on homeowners insurance consist of two parts – claims on the main dwelling (i.e. house) and claims on other structures on the property (such as a garage). Let X be the portion of a claim on the house, let Y be the portion of the same claim on the other structures.

X, Y have the joint density function

$$\begin{cases} f(x,y) & \text{if } 0 < x < 2, \quad 0 < y < 3 \\ 0 & \text{if elsewhere} \end{cases}$$

Find the probability that the total claim exceeds 3.

Solution

The 2-D region is the shaded area below (triangle ADE in Figure 12).

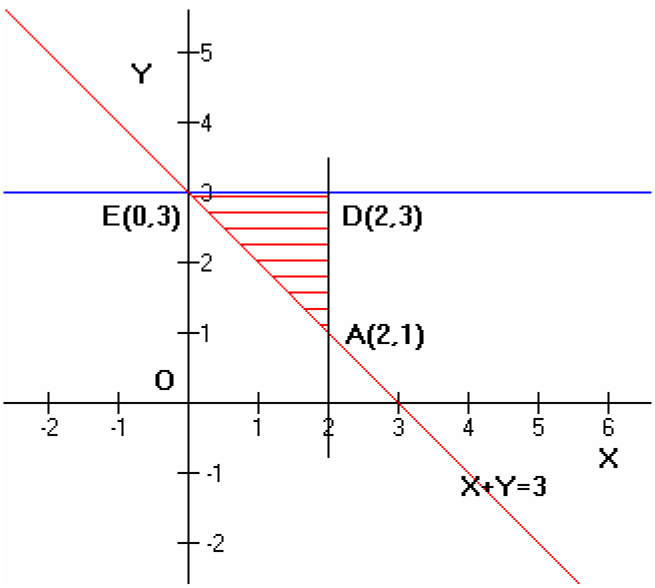


Figure 12

$$\Pr(X+Y>3)=\int_0^2 \int_{3-x}^{3} f(x,y)\,dy\,dx$$

Problem 8

Claims on homeowners insurance consist of two parts – claims on the main dwelling (i.e. house) and claims on other structures on the property (such as a garage). Let X be the portion of a claim on the house, let Y be the portion of the same claim on the other structures.

X, Y have the joint density function

$$\begin{cases} f(x,y) & \text{if } 0<x<2, \quad 0<y<3 \\ 0 & \text{if elsewhere} \end{cases}$$

Find the probability that the total claim exceeds 1 but is less than 3 and the other structures portion of the same claim is less than 1.5.

Solution

We need to find $\Pr\left[(X+Y>1)\cap(X+Y<3)\cap(Y<1.5)\right]$

The 2-D region is the shaded area (ABCDEF) in Figure 13.

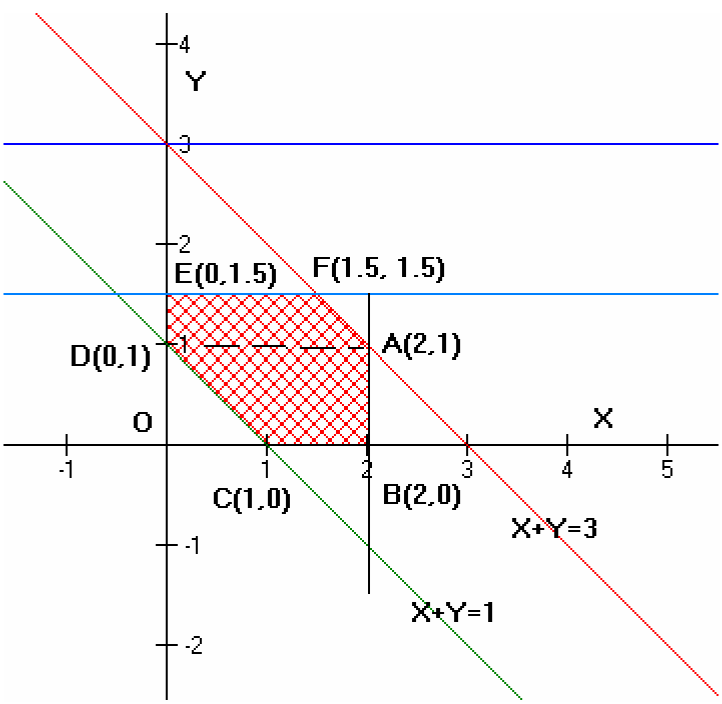


Figure 13

To do the double integration, we divide ABCDEF into ABCD and DEFA.

$$\Pr\left[(X+Y>1)\cap(X+Y<3)\cap(Y<1.5)\right]=\int_{0}^{1}\int_{1-y}^{2} f(x,y)dxdy+\int_{1}^{1.5}\int_{0}^{3-y} f(x,y)dxdy$$

Homework for you:

#5, #10, May 2000; #20, #36, #40 Nov 2000; #5, #13, #22, #24 May 2001; #28, #30 Nov 2001; #10, #20, #24 May 2003.

Chapter 29 Marginal/conditional density

How to find the marginal density if you know the joint density:

Step 1 – Draw the 2-D region for all (X, Y)
Step 2 – Set up the single integration

The X -marginal density is
$f(x) = \int_{y|x} f(x, y) dy$ (integrates over all y 's given $X = x$)

The Y -marginal density is
$f(y) = \int_{x|y} f(x, y) dx$ (integrates over all x 's given $Y = y$)

Step 3 – If the integration is on Y , draw a vertical line that cuts through the 2-D region. The Y coordinates of the two cutting points are the lower bound and higher bound of the integration.

If the integration is on X , draw a horizontal line that cuts through the 2-D region. The X coordinates of the two cutting points are the lower bound and higher bound of the integration.

You can see that the procedure for finding the marginal density is similar to the procedure for tackling problems on joint density and double integration. The only difference is that when doing double integration, you have to set up the outer integration. In contrast, the marginal density is single integration. If you take out the step related to the outer integration, the procedure for joint density problems becomes the procedure for finding the marginal density.

Formulas for the conditional density and conditional expectation

The conditional density of X given $Y = y$:

$$f_{x|y}(x|y) = \frac{f(x, y)}{f(y)}$$

The conditional expectation of X given $Y = y$:

$$E(X|Y = y) = \int_{x|y} x f_{x|y}(x|y)\, dx$$

The conditional density of Y given $X = x$:

$$f_{y|x}(y|x) = \frac{f(x, y)}{f(x)}$$

The conditional expectation of Y given $X = x$:

$$E(Y|X=x) = \int_{y|x} y\, f_{y|x}(y|x)\, dy$$

To find the conditional range $y|x$, we draw a vertical line that cuts through the 2-D region. The two cutting points are the lower and upper bounds for $y|x$. To find the conditional range $x|y$, we draw a horizontal line that cuts through the 2-D region. The two cutting points are the lower and upper bounds for $x|y$. By now you should be very good at this.

Problem 1

Random variables X and Y have the following joint distribution:

$$f(x,y) = \begin{cases} \frac{3}{8}(1-x+2y), & \text{for } 0 \le x \le y \le 2-x \\ 0, & \text{elsewhere} \end{cases}$$

Find the X-marginal density $f(x)$ and the Y-marginal density $f(y)$.

Solution

We worked on this example before when we were solving joint density related problems. Now we are finding the marginal density.

Step 1 – Draw the 2-D region for (X, Y):

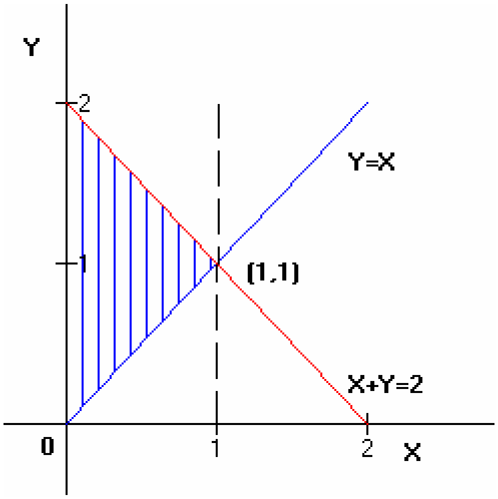


Figure 16 shaded area= $0 \le x \le y \le 2-x$

Step 2 – Set up the integration:

The X -marginal density is

$$f(x) = \int_{y|x} f(x, y)dy = \int_{y|x} \frac{3}{8}(1 - x + 2y)\, dy$$

The Y -marginal density is

$$f(y) = \int_{x|y} f(x, y)dx = \int_{x|y} \frac{3}{8}(1 - x + 2y)\, dx$$

Step 3—Determine the lower and upper bounds for integration.

If the inner integration is on Y (when finding the X -marginal):

To find the lower and upper bounds on Y, we draw a vertical line $X = x$ that cuts through the 2-D region. The two cutting points are M($x, 2 - x$) and N(x, x). So the lower bound is x (the Y coordinate of N) and the upper bound is $2 - x$ (the Y coordinate of M).

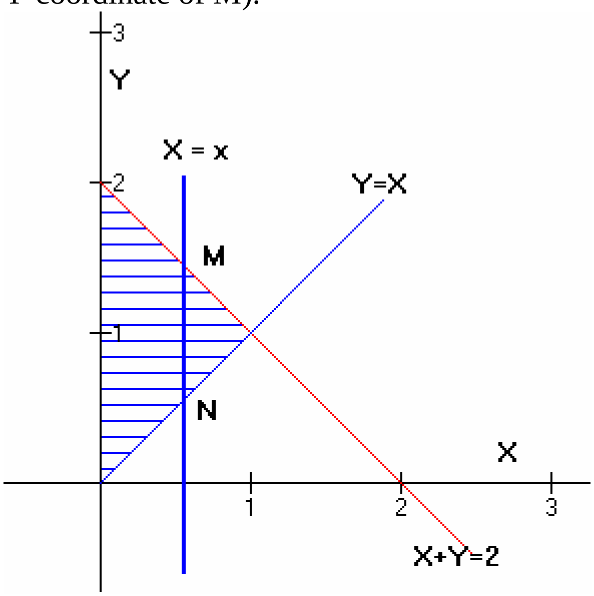


Figure 17

$$f(x) = \int_{y|x} \frac{3}{8}(1 - x + 2y)\, dy = \int_{x}^{2-x} \frac{3}{8}(1 - x + 2y)dy$$

$$\Rightarrow \quad f(x) = \int_x^{2-x} \frac{3}{8}(1 - x + 2y)dy = \frac{3}{8}(y - xy + y^2)\Big|_x^{2-x}$$

$$= \frac{3}{8}\left\{ \left[(2-x) - x(2-x) + (2-x)^2\right] - \left[x - x^2 + x^2\right] \right\}$$

$$= \frac{3}{8}(2x^2 - 8x + 6) = \frac{3}{4}(x^2 - 4x + 3) \qquad \text{for } 0 \le x \le 1$$

Because in the 2-D region, $0 \le x \le 1$, then $f(x) = \frac{3}{4}(x^2 - 4x + 3)$ is defined on $0 \le x \le 1$.

Double check: We should have $\int f(x)dx = 1$

$$\int_0^1 \frac{3}{4}(x^2 - 4x + 3)dx = \frac{3}{4}\left(\frac{1}{3}x^3 - 2x^2 + 3x\right)\Big|_0^1 = \frac{3}{4}\left(\frac{1}{3} - 2 + 3\right) = 1 \quad \text{(OK)}$$

If we do inner integration on X (when finding Y -marginal):

To determine the lower and upper bounds of X, we need to draw two horizontal lines DE and GH that cut through the 2-D region.

For the line D(0, y) and E(2 − y, y), the lower bound of X is zero (the X coordinate of D) and the upper bound is $2 - y$ (the X coordinate of E); for the line G(0, y) and H(y, y), the lower bound is zero (the X coordinate of G) and the upper bound is y (the X coordinate of H).

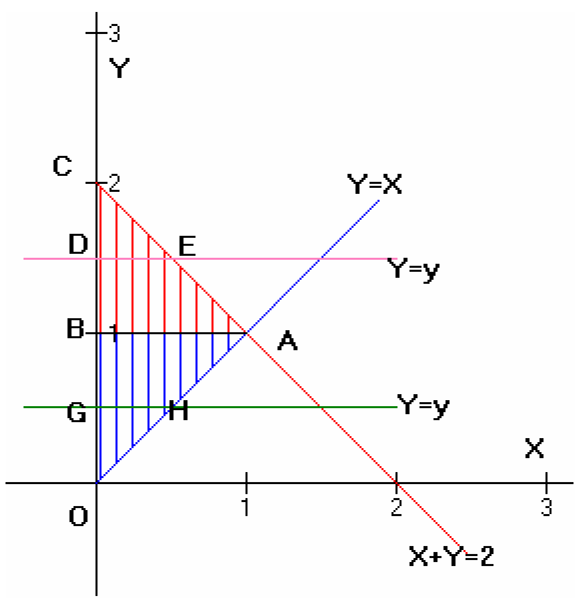


Figure 18

As a result, we need to do two separate integrations:

For $0 \le y \le 1$, the Y -marginal density is:

$$f(y) = \int_{\text{Line GH}} f(x,y)dx = \int_0^y f(x,y)dx = \int_0^y \frac{3}{8}(1-x+2y)dx$$

$$\int_0^y \frac{3}{8}(1-\underline{\underline{x}}+2y)d\underline{\underline{x}} = \frac{3}{8}\left[(1+2y)\underline{\underline{x}} - \frac{1}{2}\underline{\underline{x}}^2\right]\Bigg|_0^y = \frac{3}{8}\left[(1+2y)y - \frac{1}{2}y^2\right] = \frac{3}{8}\left(\frac{3}{2}y^2 + y\right)$$

For $1 \le y \le 2$, the Y -marginal density is:

$$f(y) = \int_{\text{Line DE}} f(x,y)dx = \int_0^{2-y} f(x,y)dx = \int_0^{2-y} \frac{3}{8}(1-x+2y)dx$$

$$\int_0^{2-y} \frac{3}{8}(1-\underline{\underline{x}}+2y)d\underline{\underline{x}} = \frac{3}{8}\left[(1+2y)\underline{\underline{x}} - \frac{1}{2}\underline{\underline{x}}^2\right]\Bigg|_0^{2-y}$$

$$= \frac{3}{8}\left[(1+2y)(2-y) - \frac{1}{2}(2-y)^2\right] = \frac{3}{8}\left(-\frac{5}{2}y^2 + 5y\right)$$

So the Y - marginal density is:

$$f(y) = \begin{cases} \dfrac{3}{8}\left(\dfrac{3}{2}y^2 + y\right) & \text{for } 0 \le y \le 1 \\[2ex] \dfrac{3}{8}\left(-\dfrac{5}{2}y^2 + 5y\right) & \text{for } 1 \le y \le 2 \end{cases}$$

Double check: We should have $\int f(y)dy = 1$

$$\int f(y)dy = \int_0^1 \frac{3}{8}\left(\frac{3}{2}y^2 + y\right)dy + \int_1^2 \frac{3}{8}\left(-\frac{5}{2}y^2 + 5y\right)dy$$

$$= \frac{3}{8}\left(\frac{1}{2}y^3 + \frac{1}{2}y^2\right)\Bigg|_0^1 + \frac{3}{8}\left(-\frac{5}{6}y^3 + \frac{5}{2}y^2\right)\Bigg|_1^2$$

$$= \frac{3}{8}\left(\frac{1}{2} + \frac{1}{2}\right) + \frac{3}{8}\left(-\frac{5}{6}\times 7 + \frac{5}{2}\times 3\right) = 1 \quad \text{(OK)}$$

Problem 2

Random variables X and Y have the following joint distribution:

$$f(x,y)=\begin{cases}\frac{3}{8}(1-x+2y), & \text{for } 0\le x\le y\le 2-x\\ 0, & \text{elsewhere}\end{cases}$$

Find $f_{x|y}(x|y)$, $f_{y|x}(y|x)$, $E(X|Y=y)$, $E(Y|X=x)$.

Solution

$$f_{x|y}(x|y)=\frac{f(x,y)}{f(y)}=\begin{cases}\dfrac{\frac{3}{8}(1-x+2y)}{\frac{3}{8}\left(\frac{3}{2}y^2+y\right)}=\dfrac{1-x+2y}{\frac{3}{2}y^2+y} & (\text{for } 0<y\le 1)\\[2ex] \dfrac{\frac{3}{8}(1-x+2y)}{\frac{3}{8}\left(-\frac{5}{2}y^2+5y\right)}=\dfrac{1-x+2y}{-\frac{5}{2}y^2+5y} & (\text{for } 1\le y\le 2)\end{cases}$$

$$f_{y|x}(y|x)=\frac{f(x,y)}{f(x)}=\frac{\frac{3}{8}(1-x+2y)}{\frac{3}{4}(x^2-4x+3)}=\frac{1-x+2y}{2(x^2-4x+3)}\quad(\text{for } 0\le x\le 1)$$

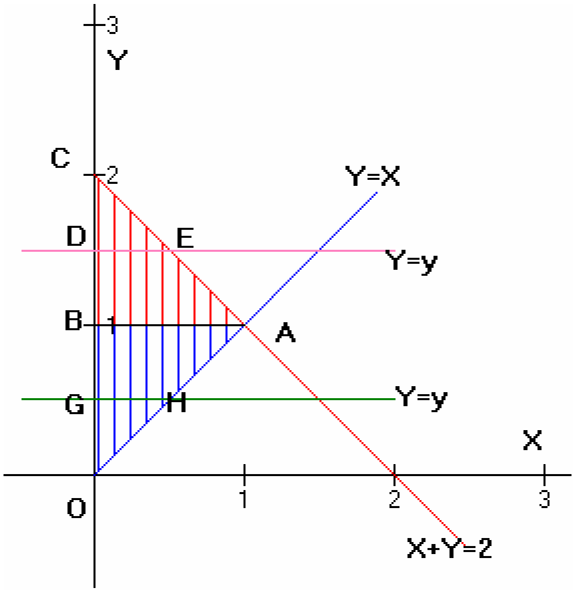


Figure 19

For $0 < y \le 1$

$$E(X|Y=y)=\int_{x|y} x\, f_{x|y}(x|y)dx=\int_{GH} x\, f_{x|y}(x|y)dx=\int_{GH} x\frac{1-x+2y}{\frac{3}{2}y^2+y}dx$$

$$=\int_0^y x\frac{1-x+2y}{\frac{3}{2}y^2+y}dx=\left(\frac{1}{\frac{3}{2}y^2+y}\right)\int_0^y x(1-x+2y)\,dx$$

$$=\left(\frac{1}{\frac{3}{2}y^2+y}\right)\int_0^y (x-x^2+2xy)\,dx=\left(\frac{1}{\frac{3}{2}y^2+y}\right)\left(\frac{1}{2}x^2-\frac{1}{3}x^3+x^2y\right)\Big|_0^y$$

$$=\left(\frac{1}{\frac{3}{2}y^2+y}\right)\left(\frac{1}{2}y^2-\frac{1}{3}y^3+y^3\right)=\left(\frac{1}{\frac{3}{2}y^2+y}\right)\left(\frac{1}{2}y^2+\frac{2}{3}y^3\right)$$

For $1 \le y \le 2$

$$E(X|Y=y)=\int_{x|y} x\, f_{x|y}(x|y)dx=\int_{DE} x\, f_{x|y}(x|y)dx=\int_{DE} x\frac{1-x+2y}{-\frac{5}{2}y^2+5y}dx$$

$$= \int_0^{2-y} x \frac{1-x+2y}{-\frac{5}{2}y^2+5y} dx = \left(\frac{1}{-\frac{5}{2}y^2+5y} \right) \int_0^{2-y} x(1-x+2y)\, dx$$

$$= \left(\frac{1}{-\frac{5}{2}y^2+5y} \right) \int_0^{2-y} (x - x^2 + 2xy)\, dx = \left(\frac{1}{-\frac{5}{2}y^2+5y} \right) \left(\frac{1}{2}x^2 - \frac{1}{3}x^3 + x^2 y \right) \Big|_0^{2-y}$$

$$= \left(\frac{1}{-\frac{5}{2}y^2+5y} \right) \left[\frac{1}{2}(2-y)^2 - \frac{1}{3}(2-y)^3 + (2-y)^2 y \right]$$

One quick check: $E(X|Y=y)$ for $0 < y \le 1$ and $E(X|Y=y)$ for $1 \le y \le 2$ should generate the identical result for $y=1$. You can verify that this is, indeed, the case; both generate the result that $E(X|Y=y)$=7/15.

Next, we find $E(Y|X=x) = \int_{y|x} y\, f_{y|x}(y|x)\, dy$

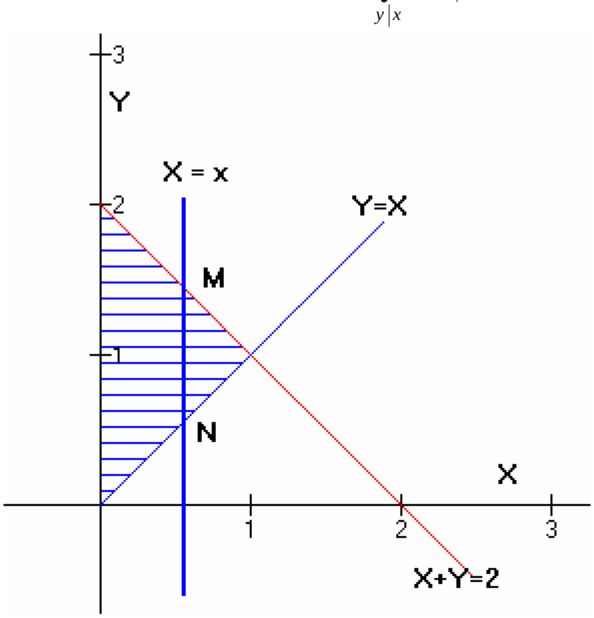


Figure 20

$$E(Y|X=x) = \int_{y|x} y\, f_{y|x}(y|x)\, dy = \int_{NM} y\, f_{y|x}(y|x)\, dy$$

$$f_{y|x}(y|x) = \frac{f(x,y)}{f(x)} = \frac{1-x+2y}{2(x^2-4x+3)} \quad (\text{for } 0 \le x \le 1)$$

$$E(Y|X=x) = \int_{NM} y\frac{1-x+2y}{2(x^2-4x+3)}dy = \int_x^{2-x} y\frac{1-x+2y}{2(x^2-4x+3)}dy$$

$$= \frac{1}{2(x^2-4x+3)}\int_x^{2-x} y(1-x+2y)\,dy$$

$$= \frac{1}{2(x^2-4x+3)}\left[\frac{1}{2}(1-x)y^2 + \frac{2}{3}y^3\right]\Bigg|_x^{2-x} = \ldots$$

(Further evaluation of $E(Y|X=x)$ is omitted)

Most likely, Exam P won't ask you to do such hard-core integrations as we did here. However, this problem does give you a good idea on how to find the marginal and conditional density.

Problem 3 (easier calculation)

Two loss random variables X and Y have the following joint distribution:

$$f(x,y) = \begin{cases} \frac{3}{8}(1-x+2y), & \text{for } 0 \le x \le y \le 2-x \\ 0, & \text{elsewhere} \end{cases}$$

Without using any of the formulas derived in the previous problem, find $\Pr(X > 0.5|Y=1)$, $E(X|Y=1)$, and $Var(X|Y=1)$.

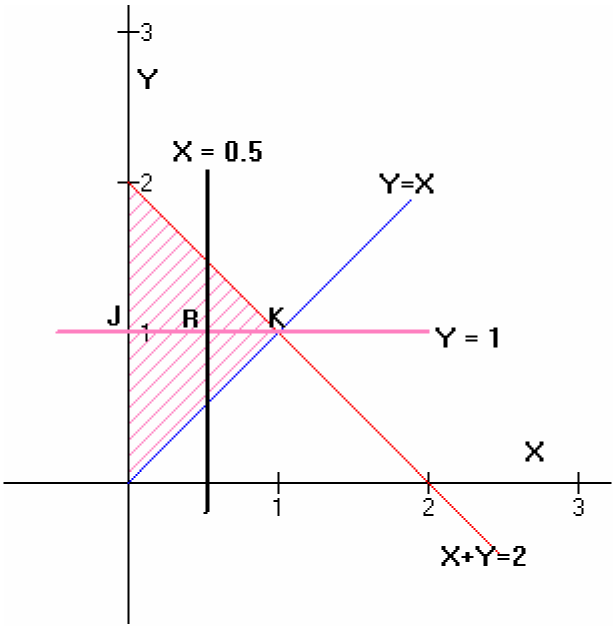


Figure 21

Solution

$$\Pr(X > 0.5 | Y = 1) = \frac{\Pr(X > 0.5, Y = 1)}{\Pr(Y = 1)}$$

$$\Pr(X > 0.5, Y = 1) = \int_{RK} f(x, y = 1)dx = \int_{0.5}^{1} f(x, y = 1)dx$$

$$f(x, y = 1) = \left[\frac{3}{8}(1 - x + 2y)\right]_{y=1} = \frac{3}{8}(3 - x)$$

$$\Pr(X > 0.5, Y = 1) = \int_{0.5}^{1} \frac{3}{8}(3 - x)dx = -\int_{2.5}^{2} \frac{3}{8}t\,dt = \frac{3}{8}\left[\frac{1}{2}t^2\right]_{2}^{2.5} = \frac{27}{64}$$

(We set $3 - x = t$ to speed up the integration)

$$\Pr(Y = 1) = \int_{JK} f(x, y = 1)dx = \int_{0}^{1} f(x, y = 1)dx = \int_{0}^{1} \frac{3}{8}(3 - x)dx$$

$$= -\int_{3}^{2} \frac{3}{8}t\,dt = \frac{3}{8}\left[\frac{1}{2}t^2\right]_{2}^{3} = \frac{15}{16}$$

$$\Pr(X > 0.5 | Y = 1) = \frac{\Pr(X > 0.5, Y = 1)}{\Pr(Y = 1)} = \frac{27}{64} \Big/ \frac{15}{16} = \frac{9}{20}$$

$$E(X|Y=1)=\int_{JK} x\frac{f(x,y=1)}{f(y=1)}dx=\int_0^1 x\frac{f(x,y=1)}{f(y=1)}dx$$

$$=\int_0^1 x\frac{\frac{3}{8}(3-x)}{\frac{15}{16}}dx=\frac{2}{5}\int_0^1 x(3-x)dx=\frac{2}{5}\int_0^1(3x-x^2)dx$$

$$=\frac{2}{5}\left[\frac{3}{2}x^2-\frac{1}{3}x^3\right]_0^1=\frac{7}{15}$$

$E(X|Y=1)=7/15$ matches the result obtained in the previous problem.

$$Var(X|Y=1)=E(X^2|Y=1)-E^2(X|Y=1)$$

$$E(X^2|Y=1)=\int_{JK} x^2\frac{f(x,y=1)}{f(y=1)}dx=\int_0^1 x^2\frac{f(x,y=1)}{f(y=1)}dx$$

$$=\int_0^1 x^2\frac{\frac{3}{8}(3-x)}{\frac{15}{16}}dx=\frac{2}{5}\int_0^1 x^2(3-x)dx=\frac{2}{5}\int_0^1(3x^2-x^3)dx$$

$$=\frac{2}{5}\left[x^3-\frac{1}{4}x^4\right]_0^1=\frac{3}{10}$$

$$Var(X|Y=1)=E(X^2|Y=1)-E^2(X|Y=1)$$

$$Var(X|Y=1)=\frac{3}{10}-\left(\frac{7}{15}\right)^2=\frac{37}{450}$$

Homework for you: #23 May 2000; #7 Nov 2000; #39 May 2001; #17,#34 Nov 2001; #28 May 2003.

Chapter 30 Transformation: CDF, PDF, and Jacobian Method

If you are given the pdf of a continuous random variable X, what is the pdf for $\mathrm{Y} = \mathrm{f}(\mathrm{X})$? If you are given $f(x, y)$, the joint pdf of X, Y, what's the joint pdf for $Z_1 = g(X, Y)$ and $Z_2 = h(X, Y)$? There are three methods:

Transformation of one variable

- CDF method (good for one-to-one and one-to-many transformation)
- PDF method (good for one-to-one transformation)

Transformation of n variables

- Jacobian method

Transformation of one variable -- CDF method.

You first calculate the CDF for the new variable. Then you differentiate the CDF to get the pdf.

Problem 1

Random variable X has the following distribution:

$$f(x) = \frac{1}{8}(x+2), \quad -2 \le x \le 2$$

Find the pdf of a random variable Y, where $Y = X^2 \quad (0 \le Y \le 4)$

Solution

Let's draw a rough diagram of $Y = X^2 \quad (0 \le Y \le 4)$. See Figure 14.

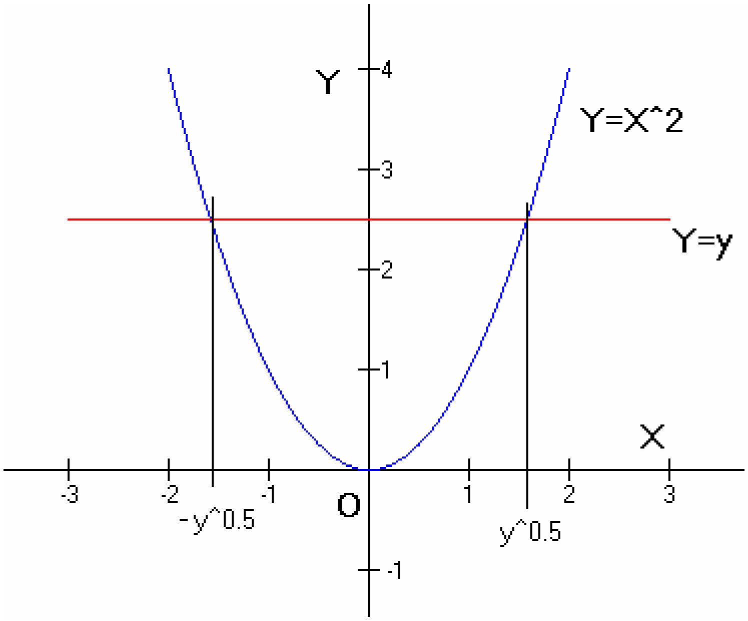


Figure 14

From Figure 14 you can see:

$$F(y) = \Pr(Y \le y) = \Pr(-\sqrt{y} \le X \le \sqrt{y})$$

We easily calculate $\Pr(-\sqrt{y} \le X \le \sqrt{y})$:

$$\Pr(-\sqrt{y} \le X \le \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f(x)dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{8}(x+2)dx = \frac{1}{8}\left(\frac{1}{2}x^2 + 2x\right)\Big|_{-\sqrt{y}}^{\sqrt{y}} = \frac{1}{2}\sqrt{y}$$

Double check:

$$F(Y=0) = \frac{1}{2}\sqrt{0} = 0 \text{ ,} F(Y=4) = \frac{1}{2}\sqrt{4} = 1 \quad \text{(OK)}$$

$$f(y) = \frac{dF(y)}{dy} = \frac{d}{dy}\left(\frac{1}{2}\sqrt{y}\right) = \frac{1}{4\sqrt{y}}, \quad (0<y \le 4)$$

Problem 2

X uniformly distributed over $[-10,10]$. $Y = |X|$. Find the pdf of Y.

Solution

X uniformly distributed over $[-10,10]$

$$\Rightarrow \quad f_X(x)=\frac{1}{20}, \quad F_X(x)=\int_{-10}^{x} f_X(s)\,ds=\int_{-10}^{x}\frac{1}{20}ds=\frac{1}{20}(x+10)$$

$$\Pr(Y \le y)=\Pr(|X| \le y)=\Pr(-y \le X \le y)=F_X(y)-F_X(-y)$$
$$=\frac{1}{20}(y+10)-\frac{1}{20}(-y+10)=\frac{y}{10} \quad \text{where } 0 \le y \le 10$$

$$\Rightarrow \quad f_Y(y)=\frac{d}{dy}F_Y(y)=\frac{1}{10} \quad \text{where } 0 \le y \le 10$$

Problem 3

A new regulation has a mandatory payment clause, which requires an insurance company to double its bodily injury claim payments to an auto insurance policyholder. Prior to this new regulation, bodily injury-related claim payments to a policyholder, X, has a pdf $f_X(x)$.

After the new regulation becomes effective, what is the probability distribution function of bodily injury claim payments to a policyholder?

Solution

Let Y =bodily injury claim payments made to the policyholder under the new law.
$Y = 2X$. We need to find the pdf $f(y)$.

$$F(y)=\Pr(Y \le y)=\Pr(2X \le y)=\Pr(X \le \frac{y}{2})=F_X\left(\frac{y}{2}\right)$$
$$f(y)=\frac{d}{dy}F_Y(y)=\frac{d}{dy}F_X\left(\frac{y}{2}\right)=\left[\frac{d}{dy}\left(\frac{y}{2}\right)\right]\frac{d}{dy}F=\frac{1}{2}f_X\left(\frac{y}{2}\right)$$

Problem 4

A machine has two parallel components backing up each other. The machine works as long as at least one of the components is working.
Each component's time until failure is independently exponentially distributed with parameters $\lambda_1=5$ and $\lambda_2=8$ respectively.

Find the probability distribution function of the machine's time until failure.

Solution
Let $T1$ and $T2$ represent each component's time until failure. Let T represent the machine's time until failure.

$T = \max(T1, T2)$

$$\Pr(T \le t) = \Pr\left[\max(T1,T2) \le t)\right] = \Pr\left[(T1 \le t) \cap (T2 \le t)\right]$$

Because $T1$ and $T2$ are independent, we have

$$\Pr\left[(T1 \le t) \cap (T2 \le t)\right] = \Pr(T1 \le t)\Pr(T2 \le t) = (1 - e^{-5t})(1 - e^{-8t})$$
$$=1 - e^{-5t} - e^{-8t} + e^{-13t}$$

$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt}(1 - e^{-5t} - e^{-8t} + e^{-13t}) = 5e^{-5t} + 8e^{-8t} - 13e^{-13t}$$

Problem 5

A machine has two components working together. The machine works only if both components are working.

Each component's time until failure is independently exponentially distributed with $\lambda_1 = 5$ and $\lambda_2 = 8$ respectively.

Find the probability distribution function of the machine's time until failure.

Solution

Let $T1$ and $T2$ represent each component's time until failure. Let T represent the machine's time until failure. $T = \min(T1, T2)$

$$\Pr(T > t) = \Pr\left[\min(T1, T2) > t)\right] = \Pr\left[(T1 > t) \cap (T2 > t)\right]$$

Because $T1$ and $T2$ are independent, we have

$$\Pr\left[(T1 > t) \cap (T2 > t)\right] = \Pr(T1 > t)\Pr(T2 > t) = (e^{-5t})(e^{-8t}) = e^{-13t}$$

Then

$$F(t) = 1 - \Pr(T > t) = 1 - e^{-13t}$$
$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt}(1 - e^{-13t}) = 13e^{-13t}$$

Problem 6

X, Y has the following joint pdf:

$$f(x, y) = x + y \text{ where } 0 \le x \le 1 \text{ and } 0 \le y \le 1$$

Let $Z = x + y$. Find $f(z)$, the pdf of Z.

Solution

First, we need to find cdf $F(z) = \Pr(Z \le z)$. Then we differentiate $F(z)$ to find $f(z)$.

Cdf is $F(z) = \Pr(Z \le z) = \Pr(x + y \le z)$. But how are we going to find $\Pr(x + y \le z)$?

Before you feel scared, let's simplify $\Pr(x + y \le z)$ by setting z to a constant such as $z = 1$. Do you know how to find $\Pr(x + y \le 1)$?
If $z = 1$, the problem is:

> X, Y has the following joint pdf:
>
> $$f(x, y) = x + y \text{ where } 0 \le x \le 1 \text{ and } 0 \le y \le 1$$
>
> Find $\Pr(x + y \le 1)$.

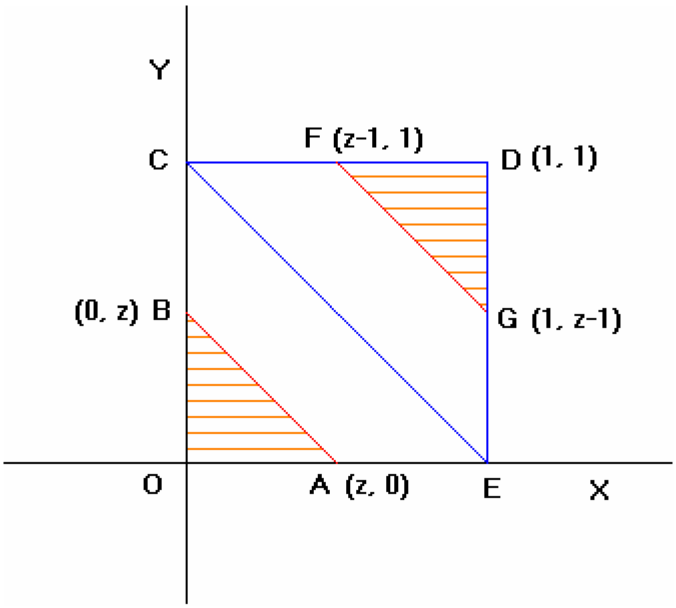
If this still doesn't ring a bell, we can translate the above simplified problem into a word problem:

> The claim amount (in the unit of \$1 million) on the house X and the claim amount on the garage Y (in the unit of \$1 million) have the following joint pdf:
>
> $$f(x, y) = x + y \text{ where } 0 \le x \le 1 \text{ and } 0 \le y \le 1$$
>
> Find the probability that the total claim amount not exceeding 1.

Now you should recognize that this problem is a joint density problem. We have a generic 4-step process described in Chapter 26 to solve this problem.

- Determine 2-D region.
- Set up the outer integration
- Set up the inner integration
- Evaluate the double integration

If you know how to solve $\Pr(x+y\le 1)$, you can apply the same 4-step process to find $\Pr(x+y\le z)$. You just treat z as a constant with the following lower and upper bonds: $0\le z=x+y\le 2$. However, we'll need to consider two situations: $0\le z\le 1$ and $1\le z\le 2$. These two situations have different 2-D regions.



Line CE is $x+y=1$.
If $0\le z\le 1$, $0\le x+y\le z$ is Area AOB.
If $1\le z\le 2$, $0\le x+y\le z$ is Area OCFGE.

If $0\le z\le 1$: $$F(z)=\Pr(X+Y\le z)=\iint_{AOB} f(x,y)\,dx\,dy=\int_0^z\int_0^{z-x}(x+y)\,dy\,dx=\frac{1}{3}z^3$$

$$\Rightarrow\quad f(z)=\frac{d}{dz}F(z)=z^2$$

If $1\le z\le 2$: $$F(z)=\Pr(X+Y\le z)=\iint_{OCFGE} f(x,y)\,dy\,dx=1-\iint_{FDG} f(x,y)\,dy\,dx$$
$$=1-\int_{z-1}^{1}\int_{z-x}^{1}(x+y)\,dy\,dx=1-\left(\frac{4}{3}-z^2+\frac{1}{3}z^3\right)=-\frac{1}{3}+z^2-\frac{1}{3}z^3$$
$$\Rightarrow\quad f(z)=\frac{d}{dz}F(z)=\frac{d}{dz}\left(-\frac{1}{3}+z^2-\frac{1}{3}z^3\right)=2z-z^2$$

$$\Rightarrow\quad f(z)=\begin{cases} z^2 & \text{for } 0\le z\le 1\\ 2z-z^2 & \text{for } 1\le z\le 2\end{cases}$$

Double check: $\int_0^2 f(z)dz$ should be one.

$$\int_0^2 f(z)dz = \int_0^1 z^2 dz + \int_1^2 \left(2z - z^2\right) dz = \left[\frac{1}{3}z^3\right]_0^1 + \left[z^2 - \frac{1}{3}z^3\right]_1^2$$
$$= \frac{1}{3} + \left(2^2 - 1\right) - \frac{1}{3}\left(2^3 - 1\right) = 1 \quad \text{(OK)}$$

General procedure for finding the pdf for $Z = g(X,Y)$, given X,Y have the joint pdf $f(x,y)$.

Step 1 – Find $F_z(z) = \Pr(Z \le z) = \Pr\left[g(X,Y) \le z\right]$.

To find $\Pr\left[g(X,Y) \le z\right]$, treat z as a constant. Now the problem becomes "Given that X,Y have the joint pdf $f(x,y)$, what's the probability that $g(X,Y) \le z$?" Use the 4-step procedure described in Chapter 26 to find $\Pr\left[g(X,Y) \le z\right]$.

Step 2 – Find $f_Z(z) = \frac{d}{dz} F_Z(z)$

Problem 7

There are two machines A and B. Let T^A =machine A's time until failure; let T^B =machine B's time until failure. T^A and T^B are independent random variables, both exponentially distributed with means of 10 and 5 respectively.

Let $X = \frac{T^A}{T^B}$. Find $f(X)$, the pdf of X $(0 < X < +\infty)$.

Solution

Because T^A, T^B are independent, their joint pdf is:

$$f(t^A, t^B) = f(t^A) f(t^B) = (1/10e^{-t^A/10})(1/5e^{-t^B/5})$$

Next, we need to find the cdf $F_X(x) = \Pr(X \le x) = \Pr\left(\frac{T^A}{T^B} \le x\right) = \Pr(T^A \le xT^B)$

Now the problem becomes "Given that T^A, T^B have the following pdf

$$(1/10e^{-t^A/10})(1/5e^{-t^B/5})$$

What's the probability that $T^A \le x T^B$ where x is a positive constant?"

Now we can use the 4-step process:

Step 1 – Determine 2-D region (shaded area in Figure 15).

Step 2 – Set up outer integration on T^B. From Figure 15, we see that the lower bound of T^B is zero; the upper bond is ∞. (You can also choose to have outer integration on T^A.)

Step 3 – Set up inner integration T^A. From Figure 15, we see that the lower bound of T^A is zero. The upper bound is $x T^B$.

Step 4 – Evaluate the double integration.

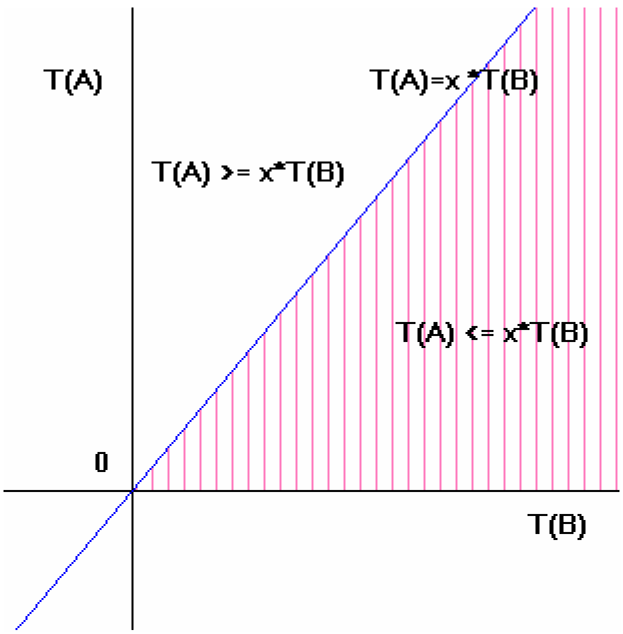


Figure 15

$$\Pr(T^A \le xT^B) = \int_0^{+\infty} \int_0^{xt^B} f(t^A, t^B)\, dt^A dt^B$$

$$= \int_0^{+\infty} \int_0^{xt^B} (1/10e^{-t^A/10})(1/5e^{-t^B/5})\, dt^A dt^B = \int_0^{+\infty} (1/5e^{-t^B/5}) \int_0^{xt^B} (1/10e^{-t^A/10})\, dt^A dt^B$$

$$= \int_0^{+\infty} (1/5e^{-t^B/5})\,(1-e^{-xt^B/10})\, dt^B$$

$$= \int_0^{+\infty} (1/5e^{-t^B/5})\,(1-e^{-xt^B/10})\,dt^B = \int_0^{+\infty} (\frac{1}{5}e^{-t^B/5})\,dt^B - \frac{1}{5}\int_0^{+\infty} e^{-\frac{x+2}{10}t^B}\,dt^B$$

$$= 1-\frac{1}{5}\frac{10}{x+2} = 1-\frac{2}{x+2}$$

Then $F(x) = 1-\dfrac{2}{x+2} \quad (0 < X < +\infty)$

Double check:

$$F(0) = 1-\frac{2}{0+2} = 0,\ F(+\infty) = 1-\frac{2}{+\infty+2} = 1 \quad \text{OK}$$

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}\left(1-\frac{2}{x+2}\right) = \frac{2}{(x+2)^2}$$

Transformation of one variable -- PDF method

Assume we already know the pdf $f_X(x)$ for X. We call "X" the old variable. Then we have a brand new variable $Y = g(X)$.

Question - How can we find $f_Y(y)$, the pdf for Y?

Answer - If the transformation is one to one and $J \neq 0$, then

$f_Y(y) = f_X(x)|J|$ Where $J = \left|\dfrac{\delta x}{\delta y}\right|$. J is called Jacobian.

Steps to find $f_Y(y)$

Step 1 Recover the old variable. From $y = g(x)$, solve for x. Make sure you get only solution. One pair means that the transformation is one to one.

Step 2 Throw the solution $x(y)$ in $f_X(x)$. You'll get $f_X[x(y)]$.

Step 3 Calculate Jacobian $J = \left|\dfrac{\delta x}{\delta y}\right|$.

Step 4 $f_Y(y) = f_X[x(y)]|J|$

Problem 8

X uniformly distributed over $[-10,10]$. $Y=|X|$. Find the pdf of Y.

Solution

X uniformly distributed over $[-10,10]$ $\Rightarrow$ $f_X(x)=\dfrac{1}{20}$,

Step 1 Recover the old variable.

$$Y=|X|, \quad \Rightarrow \quad X=\begin{cases}-Y & \text{if } X\le 0\\ Y & \text{if } X\ge 0\end{cases}$$. Notice that $Y\ge 0$.

Step 2 Throw the solution $x(y)$ in $f_X(x)$. You'll get $f_X[x(y)]$.

If $X\le 0$, $f_X[x(y)]=\dfrac{1}{20}$; If $X\ge 0$, still $f_X[x(y)]=\dfrac{1}{20}$.

Step 3 Calculate Jacobian $J=\left|\dfrac{\delta x}{\delta y}\right|$.

If $X\le 0$, $J=\left|\dfrac{\delta x}{\delta y}\right|=\left|\dfrac{\delta(-y)}{\delta y}\right|=|-1|=1$; If $X\ge 0$, $J=\left|\dfrac{\delta x}{\delta y}\right|=\left|\dfrac{\delta y}{\delta y}\right|=|1|=1$

$$\Rightarrow \quad J=\left|\frac{\delta x}{\delta y}\right|=1$$

Step 4 $f_Y(y)=f_X[x(y)]|J|$

If $X\le 0$, $f_Y(y)=f_X[x(y)]|J|=\dfrac{1}{20}$; If $X\ge 0$, $f_Y(y)=f_X[x(y)]|J|=\dfrac{1}{20}$.

$f_Y(y)$ includes both $X\le 0$ and $X\ge 0$. For example, $X=-10$ and $X=10$ both give us $Y=10$. Consequently:

$$f_Y(y)=\underbrace{f_X[x(y)]|J|}_{\text{if } X\le 0}+\underbrace{f_X[x(y)]|J|}_{\text{if } X\ge 0}=2\times\frac{1}{20}=\frac{1}{10}$$

If a transformation is one-to-many, often I find that the CDF method is easier and less prone to errors.

Problem 9 (#13, Nov 2001)

An actuary models the lifetime of a device using the random variable $Y = 10X^{0.8}$, where X is an exponential random variable with mean 1 year.

Determine the probability density function $f(y)$, for $y > 0$, of the random variable Y.

Solution

X is an exponential random variable with mean 1 year. So $f(x) = e^{-x}$.

$$Y = 10X^{0.8}, \quad \Rightarrow \quad X^{0.8} = \frac{Y}{10}, \quad X = \left(\frac{Y}{10}\right)^{\frac{1}{0.8}} = (0.1y)^{1.25} \quad \Rightarrow \quad f(x) = e^{-x} = e^{-(0.1y)^{1.25}}$$

$$J = \frac{dx}{dy} = \frac{d}{dy}(0.1y)^{1.25} = 0.1^{1.25}(1.25)y^{0.25} > 0$$

$$\Rightarrow \quad f_Y(y) = f_X[x(y)]|J| = e^{-(0.1y)^{1.25}} 0.1^{1.25}(1.25)y^{0.25} = 0.125(0.1y)^{0.25} e^{-(0.1y)^{1.25}}$$

Problem 10 (#32, Nov 2000)

The monthly profit of Company I can be modeled by a continuous random variable with density function f. Company II has a monthly profit that is twice that of Company I.

Determine the probability density function of the monthly profit of Company II.

Solution

Let X =Company I's monthly profit; Y =Company II's monthly profit.

$$Y = 2X, \Rightarrow \quad X = \frac{1}{2}Y, \quad J = \frac{dx}{dy} = \frac{1}{2} \quad \Rightarrow \quad f_Y(y) = f_X[x(y)]|J| = \frac{1}{2} f\left(\frac{y}{2}\right)$$

Transformation of n random variables

First, we consider $n = 2$, the transformation of 2 random variables. The result can be generalized for $n > 2$.

Assume we already know the joint pdf $f_{X_1,X_2}(x_1,x_2)$ of two continuous random variables X_1 and X_2. We call "X_1" and "X_2" old variables. Then we have two brand new variables $Y_1 = g_i(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$.

Question - How can we find $f_{Y_1,Y_2}(y_1,y_2)$, the joint pdf for Y_1 and Y_2?

Answer - If the transformation is one to one and $J \neq 0$, then

$$f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(x_1,x_2)|J|$$

Where $J = \det\begin{vmatrix} \dfrac{\delta x_1}{\delta y_1} & \dfrac{\delta x_1}{\delta y_2} \\ \dfrac{\delta x_2}{\delta y_1} & \dfrac{\delta x_2}{\delta y_2} \end{vmatrix} = \dfrac{\delta x_1}{\delta y_1}\dfrac{\delta x_2}{\delta y_2} - \dfrac{\delta x_1}{\delta y_2}\dfrac{\delta x_2}{\delta y_1}$. J is called Jacobian.

Steps to find $f_{Y_1,Y_2}(y_1,y_2)$

Step 1 Recover the old variables. From $y_1 = g_i(x_1,x_2)$ and $y_2 = g_2(x_1,x_2)$, solve for x_1 and x_2. Make sure you get only solution pair $\left[x_1(y_1,y_2), x_2(y_1,y_2)\right]$. One pair means that the transformation is one to one. If you get two or more solution pairs, than the transformation is not one to one. If the transformation is not one to one, you can't use the above theorem. A one-to-many transformation is beyond the scope of Exam P and you don't need to worry about it.

Step 2 Throw the solution $\left[x_1(y_1,y_2), x_2(y_1,y_2)\right]$ in $f_{X_1,X_2}(x_1,x_2)$. You'll get $f_{X_1,X_2}\left[x_1(y_1,y_2), x_2(y_1,y_2)\right]$.

Step 3 Calculate Jacobian $J = \begin{vmatrix} \dfrac{\delta x_1}{\delta y_1} & \dfrac{\delta x_1}{\delta y_2} \\ \dfrac{\delta x_2}{\delta y_1} & \dfrac{\delta x_2}{\delta y_2} \end{vmatrix} = \dfrac{\delta x_1}{\delta y_1}\dfrac{\delta x_2}{\delta y_2} - \dfrac{\delta x_1}{\delta y_2}\dfrac{\delta x_2}{\delta y_1}$.

Step 4 $f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}\left[x_1(y_1,y_2), x_2(y_1,y_2)\right]|J|$

Problem 11

X_1 and X_2 are independent identically distributed exponential random variable with mean of 1. Determine the pdf for $Y = \dfrac{X_1}{X_1 + X_2}$.

Solution

To use the theorem, we first create a fake random variable $Y_2 = X_2$. We define $Y_1 = \dfrac{X_1}{X_1 + X_2}$.

Step 1 Recover the old variables.

$$\begin{cases} Y_1 = \dfrac{X_1}{X_1 + X_2} \\ Y_2 = X_2 \end{cases}$$

$\Rightarrow \quad x_1 = \dfrac{y_1 y_2}{1 - y_1}, \quad x_2 = y_2$. We get only one solution pair; so the transformation is one-to-one. We can use the theorem.

Step 2 Throw the solution $\left[x_1(y_1, y_2), x_2(y_1, y_2)\right]$ in $f_{X_1,X_2}(x_1, x_2)$. You'll get $f_{X_1,X_2}\left[x_1(y_1, y_2), x_2(y_1, y_2)\right]$.

Let's first find the pdf for the old random variables X_1 and X_2. Because X_1 and X_2 are independent, we have:

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$$

$$f_{X_1}(x_1) = e^{-x_1}, \quad f_{X_2}(x_2) = e^{-x_2}, \qquad \Rightarrow \quad f_{X_1,X_2}(x_1, x_2) = e^{-x_1} e^{-x_2} = e^{-(x_1 + x_2)}$$

$$\Rightarrow \quad f_{X_1,X_2}(x_1, x_2) = e^{-(x_1 + x_2)} = e^{-\frac{y_1 y_2}{1 - y_1}} e^{-y_2} = e^{-\frac{y_2}{1 - y_1}}$$

Step 3 Calculate Jacobian $J = \begin{vmatrix} \dfrac{\delta x_1}{\delta y_1} & \dfrac{\delta x_1}{\delta y_2} \\ \dfrac{\delta x_2}{\delta y_1} & \dfrac{\delta x_2}{\delta y_2} \end{vmatrix} = \dfrac{\delta x_1}{\delta y_1} \dfrac{\delta x_2}{\delta y_2} - \dfrac{\delta x_1}{\delta y_2} \dfrac{\delta x_2}{\delta y_1}$.

$$\frac{\delta x_1}{\delta y_1}=\frac{d}{dy_1}\frac{y_1 y_2}{1-y_1}=\frac{y_2}{\left(1-y_1\right)^2},\quad \frac{\delta x_1}{\delta y_2}=\frac{d}{dy_2}\frac{y_1 y_2}{1-y_1}=\frac{y_1}{1-y_1},\quad \frac{\delta x_2}{\delta y_1}=\frac{dy_2}{dy_1}=0,\ \frac{\delta x_2}{\delta y_2}=1$$

$$\Rightarrow\quad J=\begin{vmatrix}\dfrac{\delta x_1}{\delta y_1} & \dfrac{\delta x_1}{\delta y_2}\\ \dfrac{\delta x_2}{\delta y_1} & \dfrac{\delta x_2}{\delta y_2}\end{vmatrix}=\begin{vmatrix}\dfrac{y_2}{\left(1-y_1\right)^2} & \dfrac{y_1}{1-y_1}\\ 0 & 1\end{vmatrix}=\frac{y_2}{\left(1-y_1\right)^2}\qquad \Rightarrow\quad \left|J\right|=\left|\frac{y_2}{\left(1-y_1\right)^2}\right|=\frac{\left|y_2\right|}{\left(1-y_1\right)^2}$$

Because X_1 and X_2 are exponential random variable, we have $X_1\geq 0$ and $X_2\geq 0$.

$$\Rightarrow\quad y_2=x_2\geq 0,\quad \left|y_2\right|=y_2,\quad \Rightarrow\quad \left|J\right|=\frac{y_2}{\left(1-y_1\right)^2}$$

Step 4 $\qquad f_{Y_1,Y_2}\left(y_1,y_2\right)=f_{X_1,X_2}\left[x_1\left(y_1,y_2\right),x_2\left(y_1,y_2\right)\right]\left|J\right|$

$$\Rightarrow\quad f_{Y_1,Y_2}\left(y_1,y_2\right)=\frac{y_2}{\left(1-y_1\right)^2}e^{-\frac{y_2}{1-y_1}}$$

Next, we eliminate y_2:

$$\Rightarrow\quad f_{Y_1}\left(y_1\right)=\int_0^{+\infty}\frac{y_2}{\left(1-y_1\right)^2}e^{-\frac{y_2}{1-y_1}}dy_2=\frac{1}{\left(1-y_1\right)^2}\int_0^{+\infty}y_2 e^{-\frac{y_2}{1-y_1}}dy_2$$

For $\int_0^{+\infty}y_2 e^{-\frac{y_2}{1-y_1}}dy_2$ to exist, we need to have $\frac{y_2}{1-y_1}\geq 0$ or $y_1<1$. This is similar to the fact that $\int_0^{+\infty}xe^{-x}dx$ exists but $\int_0^{+\infty}xe^{x}dx$ doesn't.

To find $\int_0^{+\infty}y_2 e^{-\frac{y_2}{1-y_1}}dy_2$, we set $\frac{y_2}{1-y_1}=u$. Then $y_2=u\left(1-y_1\right)$, $dy_2=\left(1-y_1\right)du$

$$\int_0^{+\infty}y_2 e^{-\frac{y_2}{1-y_1}}dy_2=\int_0^{+\infty}u\left(1-y_1\right)e^{-u}\left(1-y_1\right)du=\left(1-y_1\right)^2\int_0^{+\infty}ue^{-u}du$$

$$\int_0^{+\infty}ue^{-u}du=1$$

$$\Rightarrow \int_0^{+\infty} y_2 e^{-\frac{y_2}{1-y_1}} dy_2 = (1-y_1)^2 \int_0^{+\infty} ue^{-u} du = (1-y_1)^2$$

$$\Rightarrow \quad f_{Y_1}(y_1) = \frac{1}{(1-y_1)^2} \int_0^{+\infty} y_2 e^{-\frac{y_2}{1-y_1}} dy_2 = 1$$

So Y_1 is uniformly distributed over $[0,1]$.

Problem 12

X_1 and X_2 are independent identically exponential random variables with mean of 1. $Y = \dfrac{X_2}{X_1}$.

Find the PDF for Y.

Solution

$f_{X_1}(x_1) = e^{-x_1}$, $f_{X_2}(x_2) = e^{-x_2}$, where $x_1 > 0$ and $x_2 > 0$

$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)\, f_{X_2}(x_2) = e^{-x_1}e^{-x_2} = e^{-(x_1+x_2)}$

Let $Y_1 = X_1$ and $Y_2 = \dfrac{X_2}{X_1}$, where $Y_1 > 0$ and $Y_2 > 0$

$$\Rightarrow \quad X_1 = Y_1, \quad X_2 = Y_1 Y_2$$

$$\Rightarrow \quad f_{X_1,X_2}(x_1,x_2) = e^{-(x_1+x_2)} = e^{-y_1(1+y_2)}, \qquad \Rightarrow \quad f_{Y_1,Y_2}(y_1,y_2) = |J|\, e^{-y_1(1+y_2)}$$

$$J = \begin{vmatrix} \dfrac{\delta x_1}{\delta y_1} & \dfrac{\delta x_1}{\delta y_2} \\ \dfrac{\delta x_2}{\delta y_1} & \dfrac{\delta x_2}{\delta y_2} \end{vmatrix}, \quad \frac{\delta x_1}{\delta y_1} = 1, \quad \frac{\delta x_1}{\delta y_2} = 0, \quad \frac{\delta x_2}{\delta y_1} = \frac{\delta(y_1 y_2)}{\delta y_1} = y_2, \quad \frac{\delta x_2}{\delta y_2} = \frac{\delta(y_1 y_2)}{\delta y_2} = y_1$$

$$\Rightarrow \quad J = \begin{vmatrix} \frac{\delta x_1}{\delta y_1} & \frac{\delta x_1}{\delta y_2} \\ \frac{\delta x_2}{\delta y_1} & \frac{\delta x_2}{\delta y_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ y_2 & y_1 \end{vmatrix} = y_1, \quad |J| = |y_1| = y_1$$

$$\Rightarrow \quad f_{Y_1,Y_2}(y_1, y_2) = y_1 e^{-y_1(1+y_2)}$$

$$\Rightarrow \quad f_{Y_2}(y_2) = \int_0^{+\infty} f_{Y_1,Y_2}(y_1, y_2) dy_1 = \int_0^{+\infty} y_1 e^{-y_1(1+y_2)} dy_1 = \frac{1}{(1+y_2)^2}$$

$$\Rightarrow \quad f_Y(y) = \frac{1}{(1+y)^2} \quad \text{where } y > 0$$

Problem 13

X_1 and X_2 are independent identically exponential random variables with mean of 1.

$Y_1 = X_1 + X_2$ and $Y_2 = \frac{X_1}{X_2}$.

Find the joint PDF for Y_1 and Y_2.

Solution

$f_{X_1}(x_1) = e^{-x_1}$, $f_{X_2}(x_2) = e^{-x_2}$, where $x_1 > 0$ and $x_2 > 0$

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = e^{-x_1} e^{-x_2} = e^{-(x_1+x_2)}$$

$$Y_1 = X_1 + X_2 \text{ and } Y_2 = \frac{X_1}{X_2}. \quad \Rightarrow \quad X_1 = \frac{Y_1 Y_2}{1+Y_2}, \quad X_2 = \frac{Y_1}{1+Y_2}$$

$$\Rightarrow \quad f_{X_1,X_2}(x_1, x_2) = e^{-(x_1+x_2)} = e^{-y_1}, \quad \Rightarrow \quad f_{Y_1,Y_2}(y_1, y_2) = |J| e^{-y_1}$$

$$J = \begin{vmatrix} \frac{\delta x_1}{\delta y_1} & \frac{\delta x_1}{\delta y_2} \\ \frac{\delta x_2}{\delta y_1} & \frac{\delta x_2}{\delta y_2} \end{vmatrix}, \quad \frac{\delta x_1}{\delta y_1} = \frac{\delta}{\delta y_1}\left(\frac{y_1 y_2}{1+y_2}\right) = \frac{y_2}{1+y_2},$$

$$\frac{\delta x_1}{\delta y_2}=\frac{\delta}{\delta y_2}\left(\frac{y_1 y_2}{1+y_2}\right)=y_1\frac{\delta}{\delta y_2}\left(\frac{y_2}{1+y_2}\right)=y_1\frac{\delta}{\delta y_2}\left(1-\frac{1}{1+y_2}\right)=-y_1\frac{\delta}{\delta y_2}\left(\frac{1}{1+y_2}\right)=\frac{y_1}{\left(1+y_2\right)^2},$$

$$\frac{\delta x_2}{\delta y_1}=\frac{\delta}{\delta y_1}\left(\frac{y_1}{1+y_2}\right)=\frac{1}{1+y_2}, \qquad \frac{\delta x_2}{\delta y_2}=\frac{\delta}{\delta y_2}\left(\frac{y_1}{1+y_2}\right)=y_1\frac{\delta}{\delta y_2}\left(\frac{1}{1+y_2}\right)=-\frac{y_1}{\left(1+y_2\right)^2}$$

$$\Rightarrow \quad J=\begin{vmatrix}\frac{\delta x_1}{\delta y_1} & \frac{\delta x_1}{\delta y_2}\\ \frac{\delta x_2}{\delta y_1} & \frac{\delta x_2}{\delta y_2}\end{vmatrix}=\begin{vmatrix}\frac{y_2}{1+y_2} & \frac{y_1}{\left(1+y_2\right)^2}\\ \frac{1}{1+y_2} & -\frac{y_1}{\left(1+y_2\right)^2}\end{vmatrix}=-\frac{y_1}{\left(1+y_2\right)^2},$$

$$\Rightarrow \quad |J|=\left|-\frac{y_1}{\left(1+y_2\right)^2}\right|=\frac{y_1}{\left(1+y_2\right)^2}$$

$$\Rightarrow \quad f_{Y_1,Y_2}\left(y_1,y_2\right)=|J|e^{-y_1}=\frac{y_1}{\left(1+y_2\right)^2}e^{-y_1} \quad \text{where } y_1>0 \text{ and } y_2>0$$

Homework for you: #4 May 2000; #32, Nov 2000; 26, May 2001; #13, #37 Nov 2001; #23, May 2003.

Chapter 31 Univariate & joint order statistics

Let X_1, X_2, …, X_n represent n independent identically distributed random variables with a common pdf $f_X(x)$ and a common cdf $F_X(x)$. We assume X is a continuous random variable.

Next, we sort X_1, $X2$, …, X_n by ascending order:

$$X_{(1)} \le X_{(2)} \le X_{(3)} \le ... \le X_{(n)}$$

$$X_{(1)} = \min(X_1, X_2, ..., X_n)$$

$X_{(1)}$ is called the 1st order statistics of X_1, X_2, …, X_n.

$$X_{(2)} = \min\left[\{X_1, X_2, ..., X_n\} - \{X_{(1)}\}\right]$$

$X_{(2)}$ is the 2nd smallest number. To find $X_{(2)}$, we simply remove $X_{(1)}$ from the original set $(X_1, X_2, ..., X_n)$ and find the minimum of the remaining set.
$X_{(2)}$ is called the 2nd order statistics of X_1, X_2, …, X_n.

$$X_{(3)} = \min\left[\{X_1, X_2, ..., X_n\} - \{X_{(1)}, X_{(2)}\}\right]$$

$X_{(3)}$ is the 3rd smallest number. To find $X_{(3)}$, we simply remove $X_{(1)}$ and $X_{(2)}$ from the original set $(X_1, X_2, ..., X_n)$ and find the minimum of the remaining set. $X_{(3)}$ is called the 3rd order statistics of X_1, X_2, …, X_n.

……

$$X_{(k)} = \min\left[\{X_1, X_2, ..., X_n\} - \{X_{(1)}, X_{(2)}, ..., X_{(k-1)}\}\right]$$

$X_{(k)}$ is the *k*-th smallest number. To find $X_{(k)}$, we simply remove $X_{(1)}, X_{(2)}, …, X_{(k-1)}$ from the original set $(X_1, X_2, ..., X_n)$ and find the minimum of the remaining set. $X_{(k)}$ is called the *k*-th order statistics of X_1, X_2, …, X_n.

..

Keep doing this until

$$X_{(n)} = \min\left[\{X_1, X_2, ..., X_n\} - \{X_{(1)}, X_{(2)}, ..., X_{(n-1)}\}\right] = \max(X_1, X_2, ..., X_n)$$

$X_{(n)}$ is called the *n*-th order statistics of X_1, X_2, …, X_n.

Example. 5 random samples are drawn from a population and the values of these samples are: 5, 2, 9, 20, and 7. We sort these values into 2, 5, 7, 9, 20. So the first order is 2. The 2nd order is 5. The 3rd order is 7. The fourth order is 9. And finally, the 5th order is 20.

Now we know what order statistics is. Question – given n independent identically distributed random variables X_1, X_2, …, X_n with a common pdf $f(x)$ and a common cdf $F(X)$, what's the probability distribution (pdf) of the 1st order $X_{(1)}$? Pdf of the 2nd order $X_{(2)}$? …Pdf of the *k*-th order $X_{(k)}$? …Pdf of the *n*-th order $X_{(n)}$?

Most likely, Exam P will ask you to find the pdf of $X_{(1)}$ or $X_{(n)}$. However, we'll derive a generic pdf formula for $X_{(k)}$.

Method 1 – good only for finding the pdf of $X_{(1)}$ and $X_{(n)}$

$$X_{(1)} = \min(X_1, X_2, ..., X_n)$$
$$\Rightarrow \quad P\left[X_{(1)} \le x\right] = 1 - P\left[X_{(1)} > x\right] = 1 - P\left[\min(X_1, X_2, ..., X_n) > x\right]$$

$$P\left[\min(X_1, X_2, ..., X_n) > x\right] = P\left[(X_1 > x) \cap (X_2 > x) ... \cap (X_n > x)\right]$$

X_1, $X2$, …, X_n are independent,
$$\Rightarrow \quad P\left[(X_1 > x) \cap (X_2 > x) ... \cap (X_n > x)\right] = P(X_1 > x) P(X_2 > x) ... P(X_n > x)$$

X_1, $X2$, …, X_n are identically distributed with a common cdf $F_X(x)$
$$\Rightarrow \quad P(X_1 > x) = P(X_2 > x) = ... = P(X_n > x) = 1 - F_X(x)$$
$$\Rightarrow \quad P(X_1 > x) P(X_2 > x) ... P(X_n > x) = \left[1 - F_X(x)\right]^n$$
$$\Rightarrow \quad F_{X_{(1)}}(x) = P\left[X_{(1)} \le x\right] = 1 - P\left[(X_1 > x) \cap (X_2 > x) ... \cap (X_n > x)\right] = 1 - \left[1 - F_X(x)\right]^n$$

$$\Rightarrow \quad f_{X_{(1)}}(x) = \frac{d}{dx} F_{X_{(1)}}(x) = \frac{d}{dx}\left\{1 - \left[1 - F_X(x)\right]^n\right\} = n\left[1 - F_X(x)\right]^{n-1} \frac{d}{dx} F_X(x)$$
$$= n\left[1 - F_X(x)\right]^{n-1} f_X(x)$$

$$\boxed{\Rightarrow \quad f_{X_{(1)}}(x) = n\left[1 - F_X(x)\right]^{n-1} f_X(x)}$$

$$X_{(n)} = \max(X_1, X_2, ..., X_n)$$

$$\Rightarrow \quad P\left[X_{(n)} \leq x\right] = P\left[\max\left(X_1, X_2, ..., X_n\right) \leq x\right] = P\left[\left(X_1 \leq x\right) \cap \left(X_2 \leq x\right) ... \cap \left(X_n \leq x\right)\right]$$

X_1, $X2$, …, X_n are independent,

$$\Rightarrow \quad P\left[\left(X_1 \leq x\right) \cap \left(X_2 \leq x\right) ... \cap \left(X_n \leq x\right)\right] = P\left(X_1 \leq x\right) P\left(X_2 \leq x\right) ... P\left(X_n \leq x\right)$$

X_1, $X2$, …, X_n are identically distributed with a common cdf $F_X(x)$

$$\Rightarrow \quad P(X_1 \leq x) = P(X_2 \leq x) = ... = P(X_n \leq x) = F_X(x)$$

$$\Rightarrow \quad P\left(X_1 \leq x\right) P\left(X_2 \leq x\right) ... P\left(X_n \leq x\right) = \left[F_X\left(x\right)\right]^n$$

$$\Rightarrow \quad F_{X_{(n)}}\left(x\right) = P\left[X_{(n)} \leq x\right] = \left[F_X\left(x\right)\right]^n$$

$$\Rightarrow \quad f_{X_{(n)}}\left(x\right) = \frac{d}{dx} F_{X_{(n)}}\left(x\right) = \frac{d}{dx}\left[F_X\left(x\right)\right]^n = n\left[F_X\left(x\right)\right]^{n-1} \frac{d}{dx} F_X\left(x\right)$$

$$= n\left[F_X\left(x\right)\right]^{n-1} f_X\left(x\right)$$

$$\boxed{\Rightarrow \quad f_{X_{(n)}}\left(x\right) = n\left[F_X\left(x\right)\right]^{n-1} f_X\left(x\right)}$$

Method 2

$$X_{(1)} = \min(X_1, X_2, ..., X_n)$$

$$f_{X_{(1)}}(x)dx = P[x \leq \min(X_1, X_2, ..., X_n) \leq x + dx\ ]$$

$$= P\left[\left(x \leq \text{one of the } X\text{'s } \leq x + dx\right) \cap \left(\text{all the other } X\text{'s} > x + dx\right)\right]$$

You might wonder why I didn't write

$$f_{X_{(1)}}\left(x\right) = P\left[x \leq \min\left(X_1, X_2, ..., X_n\right) \leq x + dx\ \right]$$

The above expression is wrong. As explained before, for a continuous random variable X, the pdf $f_X\left(x\right)$ is not real probability. As a matter of fact, we don't require $f_X\left(x\right) \leq 1$; $f_X\left(x\right)$ can approach infinity. So $f_X\left(x\right)$ is not a real probability. As a result, we can't write

$$f_{X_{(1)}}\left(x\right) = P\left[x \leq \min\left(X_1, X_2, ..., X_n\right) \leq x + dx\ \right]$$

However, $f_X(x)dx$ is a genuine probability:

$$f_X(x)dx = P(x \le X < x+dx)$$

This is why I wrote:

$$f_{X_{(1)}}(x)dx = P[x \le \min(X_1, X_2, ..., X_n) \le x+dx\ ]$$
$$= P\left[(x \le \text{one of the } X\text{'s } \le x+dx) \cap (\text{all the other } X\text{'s} > x+dx)\right]$$

Let's continue.

$$P(x \le \text{one of the } X\text{'s } \le x+dx) = n\ f_X(x)dx$$
$$P(\text{all the other } X\text{'s} > x+dx) = P(\text{all the other } X\text{'s} > x) = \left[1-F_X(x)\right]^{n-1} \quad (dx \text{ is tiny})$$

$$\Rightarrow \quad f_{X_{(1)}}(x)dx = n\ f_X(x)\left[1-F_X(x)\right]^{n-1} dx$$
$$\Rightarrow \quad f_{X_{(1)}}(x) = n\ f_X(x)\left[1-F_X(x)\right]^{n-1}$$

$$X_{(n)} = \max(X_1, X_2, ..., X_n)$$

$$f_{X_{(n1)}}(x)dx = P[x \le \max(X_1, X_2, ..., X_n) \le x+dx\ ]$$
$$= P\left[(x \le \text{one of the } X\text{'s } \le x+dx) \cap (\text{all the other } X\text{'s} \le x)\right]$$

$$P\left[x \le \text{one of the } X\text{'s } \le x+dx\right] = n\ f_X(x)dx, \quad P\left[\text{all the other } X\text{'s} \le x\right] = \left[F_X(x)\right]^{n-1}$$

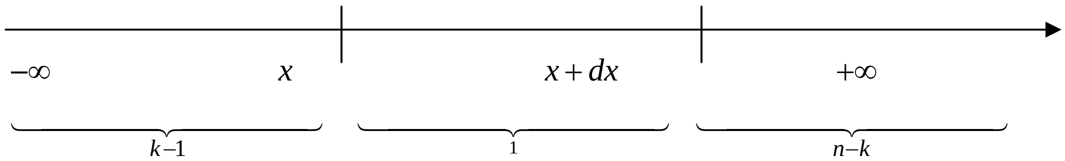
$$\Rightarrow \quad f_{X_{(n)}}(x)dx = n\ f_X(x)\left[F_X(x)\right]^{n-1} dx$$

$$\Rightarrow \quad f_{X_{(n)}}(x) = n\ f_X(x)\left[F_X(x)\right]^{n-1}$$

Extending the same logic to $X_{(k)}$:

$$f_{X_{(k)}}(x)dx$$
$$= P\{x \le \text{one of the } X\text{'s } \le x+dx \quad \cap$$
$$\text{of the remaining}(n\text{-}1)\ \ X\text{'s},\ (k-1)\ X\text{'s} \ < x,\ \ (n\text{-}1)-(k-1)=(n-k)\ \ X\text{'s} \ > x+dx\}$$

See the diagram below.



$$P\left[x \le \text{one of the } X\text{'s} \le x+dx\right] = n\ f_X(x)dx$$

$$P\left\{\text{of the remaining} (n\text{-}1)\ \ X\text{'s},\ (k-1)\ X\text{'s}\ < x,\ \ (n\text{-}1)-(k-1)=(n-k)\ \ X\text{'s}\ > x\right\}$$

$$= C_{n-1}^{k-1}\left[F_X(x)\right]^{k-1}\left[1-F_X(x)\right]^{n-k} \quad \text{(binomial distribution)}$$

$$\Rightarrow \quad f_{X_{(k)}}(x)dx = n\ f_X(x)\,C_{n-1}^{k-1}\left[F_X(x)\right]^{k-1}\left[1-F_X(x)\right]^{n-k} dx$$

$$\Rightarrow \quad f_{X_{(k)}}(x) = n\ f_X(x)\,C_{n-1}^{k-1}\left[F_X(x)\right]^{k-1}\left[1-F_X(x)\right]^{n-k}$$

Alternatively, we can derive $f_{X_{(k)}}(x)$ as follows:

$$P\left[x \le\ X \le x+dx\right] = f_X(x)dx,\ \ P(X < x) = F_X(x),\ \ P(X > x) = 1-F_X(x)$$

We have:

$1\ \ X$ falls between $(x,\, x+dx)$,
$k-1\ \ X\text{'s}$ fall between $(-\infty,\, x)$,
$n-k\ \ X\text{'s}$ fall between $(x,\, \infty)$,

If we don't worry about permutations, the pdf should be:

$$f_X(x)\left[F_X(x)\right]^{k-1}\left[1-F_X(x)\right]^{n-k}$$

Because we don't know which X falls between $(x,\, x+dx)$, which $(k-1)\ \ X\text{'s}$ fall between $(-\infty,\, x)$, and which $(n-k)\ \ X\text{'s}$ fall between $(x,\, \infty)$, the # of permutations is:

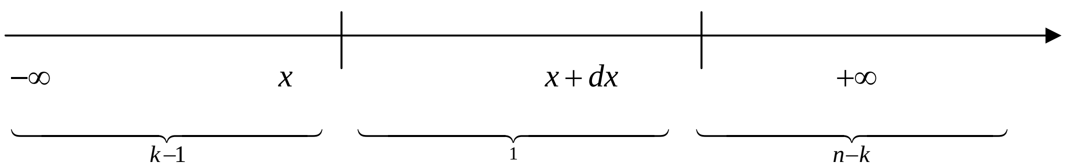
$$\frac{n!}{1!(k-1)!\left[n-1-(k-1)\right]!} = n\frac{(n-1)!}{(k-1)!\left[(n-1)-(k-1)\right]!} = n\,C_{n-1}^{k-1}$$

$$\Rightarrow \quad f_{X_{(k)}}(x) = \underbrace{\frac{n!}{1!(k-1)!\left[n-1-(k-1)\right]!}}_{\#\text{ of permutations}} \underbrace{f_X(x)}_{1\ X\ \in (x,\,x+dx)} \underbrace{\left[F_X(x)\right]^{k-1}}_{(k\text{-}1)\ X's \in (-\infty,\,x)} \underbrace{\left[1-F_X(x)\right]^{n-k}}_{(n-k)\ X's \in (x+dx,\,\infty)}$$

The above formula is the same as

$$f_{X_{(k)}}(x) = n\ f_X(x)\, C_{n-1}^{k-1} \left[F_X(x)\right]^{k-1} \left[1-F_X(x)\right]^{n-k}$$

If you can't remember $f_{X_{(k)}}(x) = n\ f_X(x)\, C_{n-1}^{k-1} \left[F_X(x)\right]^{k-1} \left[1-F_X(x)\right]^{n-k}$, just draw the following diagram:



Then write

$$f_{X_{(k)}}(x) = \underbrace{\frac{n!}{(k-1)!1!\left[n-1-(k-1)\right]!}}_{\#\text{ of permutations}} \underbrace{\left[F_X(x)\right]^{k-1}}_{(k\text{-}1)\ X's \in (-\infty,\,x)} \underbrace{f_X(x)}_{1\ X\ \in (x,\,x+dx)} \underbrace{\left[1-F_X(x)\right]^{n-k}}_{(n-k)\ X's \in (x+dx,\,\infty)}$$

Order statistics has its applications in the real world. For the purpose of passing Exam P, however, I recommend that you don't worry about how order statistics is used.

Finally, please note that $X_{(1)}$ and $X_{(n)}$ can be different from the minimum and the maximum in transformation. In Chapter 30, we calculated the pdf for the minimum and the maximum of two independent random variables that are <u>not identically distributed</u>. In contrast, $X_{(1)}$ and $X_{(n)}$ refer to the minimum and maximum of several independent <u>identically distributed</u> random variables. If several independent random variables are NOT identically distributed (such as in Chapter 30), you can't use $f_{X_{(1)}}(x) = n\left[1-F_X(x)\right]^{n-1} f_X(x)$ or $f_{X_{(n)}}(x) = n\left[F_X(x)\right]^{n-1} f_X(x)$ to find the pdf for the minimum or maximum.

Problem 1

A system has 5 duplicate components. The system works as long as at least one component works. The system fails if all components fail. The life time of each component follows a gamma distribution with parameters $n = 2$ and λ.

Find the probability distribution of the life time of the system.

Solution

Let X_1, X_2, X_3, X_4, X_5 represent the life time of each of the 5 components. Let Y represent the life time of the whole system.

Let $Y = \max(X_1, X_2, X_3, X_4, X_5)$

X_1, X_2, X_3, X_4, X_5 are independent and identically distributed with the following common pdf:

$$f_X(x) = \lambda^2 x e^{-\lambda x} \quad \text{(gamma pdf with parameters } n = 2 \text{ and } \lambda\text{)}$$

The common cdf is:

$$F_X(x) = \int_0^x \lambda^2 t e^{-\lambda t} dt = 1 - e^{-\lambda x} - \lambda x e^{-\lambda x}$$

$$\Rightarrow \quad f_Y(y) = 5 f_X(x)\left[F_X(y)\right]^4 = 5\lambda^2 x e^{-\lambda x}\left[1 - e^{-\lambda x} - \lambda x e^{-\lambda x}\right]^4$$

Problem 2

A system has 5 components working in parallel. The system works if all of the components work. The system fails if at least one component fails. The life time of each component follows a gamma distribution with parameters $n = 2$ and λ.

Find the probability distribution of the life time of the system.

Solution

Let X_1, X_2, X_3, X_4, X_5 represent the life time of each of the 5 components.

Let $Y = \min(X_1, X_2, X_3, X_4, X_5)$

This time, we are looking for the distribution of the 1st order (minimum) statistics.

Applying the formula, we have:

$$f_Y(y) = n\left[1 - F_X(x)\right]^{n-1} f_X(x) = 5\left[1 - \left(1 - e^{-\lambda x} - \lambda x e^{-\lambda x}\right)\right]^{5-1} \lambda^2 x e^{-\lambda x}$$
$$= 5\left[1 - \left(1 - e^{-\lambda x} - \lambda x e^{-\lambda x}\right)\right]^{5-1} \lambda^2 x e^{-\lambda x}$$

Problem 3

X_1, X_2, X_3 are three independent identically distributed continuous random variables with the following common pdf:

$$f(x) = \frac{8}{x^3} \quad \text{where } x \geq 2$$

Calculate $E\left[X_{(1)}\right]$, $E\left[X_{(2)}\right]$, and $E\left[X_{(3)}\right]$.

Solution

$$f_{X_{(1)}}(x) = 3 f_X(x)\left[1 - F_X(x)\right]^2$$

$$F_X(x) = \int_2^x f_x(t)dt = \int_2^x \frac{8}{t^3}dt = 1 - \frac{4}{x^2}$$

$$\Rightarrow \quad f_{X_{(1)}}(x) = 3 f_X(x)\left[1 - F_X(x)\right]^2 = 3\left(\frac{8}{x^3}\right)\left[1 - \left(1 - \frac{4}{x^2}\right)\right]^2 = \frac{384}{x^7} \quad \text{where } x \geq 2$$

$$E\left[X_{(1)}\right] = \int_2^\infty x f_{X_{(1)}}(x)\,dx = \int_2^\infty x \frac{384}{x^7}\,dx = \frac{12}{5}$$

$$f_{X_{(3)}}(x) = 3 f_X(x)\left[F_X(x)\right]^2 = 3\left(\frac{8}{x^3}\right)\left(1 - \frac{4}{x^2}\right)^2$$

$$E\left[X_{(3)}\right] = \int_2^\infty x f_{X_{(3)}}(x)\,dx = 3\int_2^\infty x\left(\frac{8}{x^3}\right)\left(1 - \frac{4}{x^2}\right)^2 dx = \frac{32}{5}$$

$$f_{X_{(2)}}(x) = 3\ f_X(x)\, C_{3-1}^{2-1}\left[F_X(x)\right]^{2-1}\left[1 - F_X(x)\right]^{3-2}$$

$$= 3\left(\frac{8}{x^3}\right)\ C_2^1\left(1 - \frac{4}{x^2}\right)\left(\frac{4}{x^2}\right) = 6\left(\frac{8}{x^3}\right)\left(1 - \frac{4}{x^2}\right)\left(\frac{4}{x^2}\right)$$

$$E\left[X_{(2)}\right]=\int_{2}^{\infty} x f_{X_{(2)}}(x)\,dx=\int_{2}^{\infty} x(6)\left(\frac{8}{x^3}\right)\left(1-\frac{4}{x^2}\right)\left(\frac{4}{x^2}\right)dx=\frac{16}{5}$$

Problem 4 (CAS Exam 3 #25, Spring 2005, modified)

Samples are selected from a uniform distribution on [0, 10].

Determine the expected value of the 4th order statistic for a sample of size five.

Solution

$$f_{X_{(k)}}(x)=n\ f_X(x)\,C_{n-1}^{k-1}\left[F_X(x)\right]^{k-1}\left[1-F_X(x)\right]^{n-k}$$

We have $n=5$ and $k=4$.

$$\Rightarrow\quad f_{X_{(4)}}(x)=5\ f_X(x)\,C_{5-1}^{4-1}\left[F_X(x)\right]^{4-1}\left[1-F_X(x)\right]^{5-4}$$

$$=5\ f_X(x)\,C_4^3\left[F_X(x)\right]^3\left[1-F_X(x)\right]$$

$$f_X(x)=\frac{1}{10},\quad F_X(x)=\frac{x}{10}$$

$$\Rightarrow\quad f_{X_{(4)}}(x)=5\ f_X(x)\,C_4^3\left[F_X(x)\right]^3\left[1-F_X(x)\right]$$

$$=5\left(\frac{1}{10}\right)C_4^3\left[\frac{x}{10}\right]^3\left[1-\frac{x}{10}\right]$$

$$=5\left(\frac{1}{10}\right)(4)\left[\frac{x}{10}\right]^3\left(1-\frac{x}{10}\right)=2\left(\frac{x}{10}\right)^3\left(1-\frac{x}{10}\right)$$

$$E\left[X_{(4)}\right]=\int_{0}^{10} x f_{X_{(4)}}(x)\,dx=\int_{0}^{10} x(2)\left(\frac{x}{10}\right)^3\left(1-\frac{x}{10}\right)dx=\frac{20}{3}$$

Homework for you: redo all the problems in this chapter.

****
****

The following is an advanced topic of order statistics: joint pdf for order statistics. Not sure whether SOA will test it. If you don' want to learn about this, just skip it.

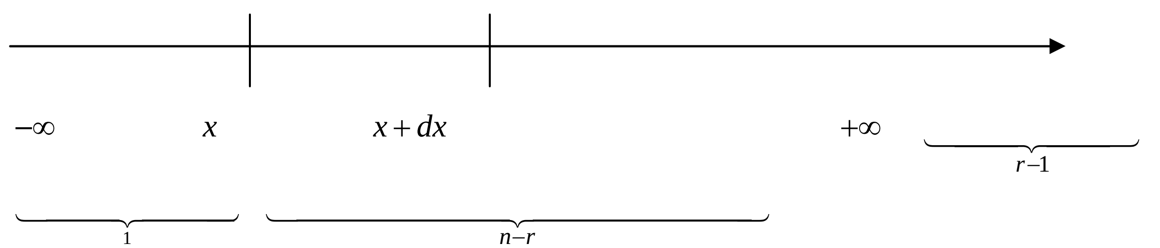
We'll answer the following two questions:

- What's the joint pdf for $X_{(r)}$ and $X_{(s)}$?
- What's the joint pdf for $X_{(1)}$, $X_{(2)}$, $X_{(3)}$, …., $X_{(n)}$?

First, we'll find the joint pdf for $X_{(r)}$ and $X_{(s)}$. We'll consider $r < s$. Our goal is to find $f_{X_{(r)}, Y_{(s)}}(x, y)$. Please note that if $r < s$, then $X_{(r)} \le X_{(s)}$. So we have $f_{X_{(r)}, Y_{(s)}}(x, y) = 0$ if $x > y$.

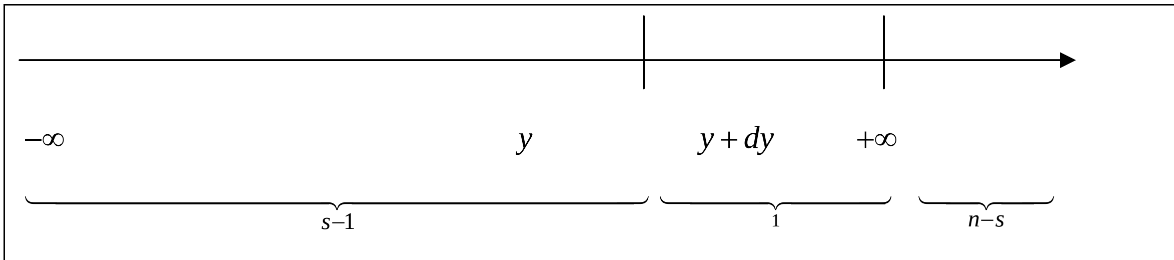
Let's consider $P\left[x < X_{(r)} < x + dx, \quad y < X_{(s)} < y + dy\right]$ where $y > x$. We'll draw three diagrams.

Diagram 1 $x < X_{(r)} < x + dx$



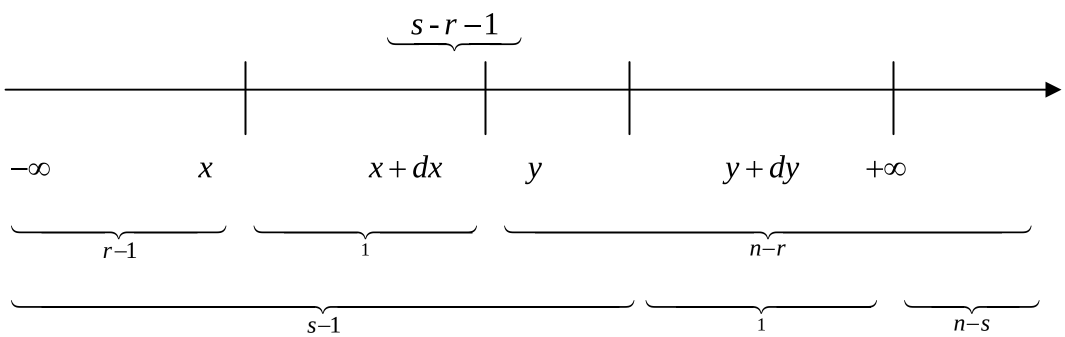
To have $x < X_{(r)} < x + dx$, we need to have one X in $(x, x + dx)$, $(r-1)$ X's in $(-\infty, x)$, and $(n-r)$ X's in $(x, +\infty)$.

Diagram 2 $y < X_{(s)} < y + dy$



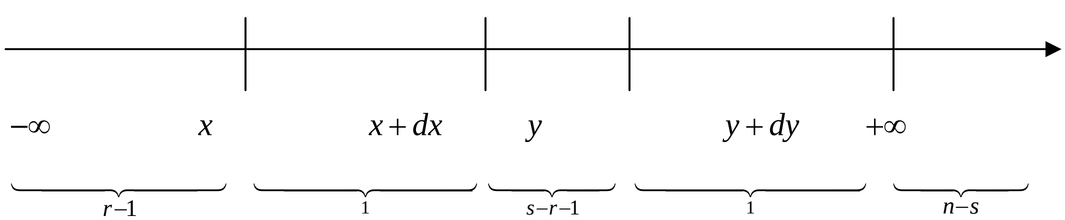
To have $y < X_{(s)} < y + dy$, we need to have one X in $(y, y+dy)$, $(s-1)$ X's in $(-\infty, y)$, and $(n-s)$ X's in $(y, +\infty)$.

Diagram 3 $x < X_{(r)} < x + dx$ and $y < X_{(s)} < y + dy$ (where $y \geq x$) -- We combine the above two diagrams into one.



To have $x < X_{(r)} < x + dx$ and $y < X_{(s)} < y + dy$ (where $y \geq x$), we need to have $(r-1)$ X's in $(-\infty, x)$, one X in $(x, x+dx)$, $(s-r-1)$ X's in $(x+dx, y)$, one X in $(y, y+dy)$, and $(n-r)$ X's in $(x, +\infty)$.

Diagram 3 can be simplified as follows:



$$f_{X_{(r)}, X_{(s)}}(x, y)\, dx\, dy = P\left[x < X_{(r)} < x + dx, \quad y < X_{(s)} < y + dy\right]$$

$$= A \times \underbrace{\left[F_X(x)\right]^{r-1}}_{(r\text{-}1)\ X's \in (-\infty, x)} \underbrace{f_X(x)\, dx}_{1\ X\ \in (x, x+dx)} \underbrace{\left[F_X(y) - F_X(x)\right]^{s-r-1}}_{(s-r-1)\ X's \in (x+dx, y)} \underbrace{f_X(y)\, \mathrm{d}y}_{1\ X\ \in (y, y+dy)} \underbrace{\left[1 - F_X(y)\right]^{n-s}}_{(n-s)\ X's \in (y+dy, \infty)}$$

Where $A = \underbrace{\dfrac{n!}{(r-1)!1!(s-r-1)!1!(n-s)!}}_{\text{\# of permutations}} = \dfrac{n!}{(r-1)!(s-r-1)!(n-s)!}$

Formula:

$$f_{X_{(r)},X_{(s)}}(x,y)=$$
$$\frac{n!}{(r-1)!(s-r-1)!(n-s)!}\left[F_X(x)\right]^{r-1}\left[F_X(y)-F_X(x)\right]^{s-r-1}\left[1-F_X(y)\right]^{n-s} f_X(x)\,f_X(y)$$

Where $s \geq r$ and $y \geq x$

Problem 5

X_1, X_2, X_3, …, X_n are independent random variables uniformly distributed over $[0,1]$. Find the joint pdf for $X_{(1)}$ and $X_{(n)}$.

Solution

Here $r=1$, $s=n$. Using the memorized formula, we have:

$$f_{X_{(1)},X_{(n)}}(x,y)$$

$$=\frac{n!}{(r-1)!(s-r-1)!(n-s)!}\left[F_X(x)\right]^{r-1}\left[F_X(y)-F_X(x)\right]^{s-r-1}\left[1-F_X(y)\right]^{n-s} f_X(x)\,f_X(y)$$

$$=\frac{n!}{(1-1)!(n-1-1)!(n-n)!}\left[F_X(x)\right]^{1-1}\left[F_X(y)-F_X(x)\right]^{n-1-1}\left[1-F_X(y)\right]^{n-n} f_X(x)\,f_X(y)$$

$$=\frac{n!}{(n-2)!}\left[F_X(y)-F_X(x)\right]^{n-2} f_X(x)\,f_X(y)$$

$$=n(n-1)\left[F_X(y)-F_X(x)\right]^{n-2} f_X(x)\,f_X(y)$$

X is uniform over $[0,1]$. So $f_X(x)=f_X(y)=1$, $F_X(x)=\int_0^x f(t)dt=\int_0^x dt=x$, $F_X(y)=y$.

$$\Rightarrow \quad f_{X_{(1)},X_{(n)}}(x,y)=n(n-1)(y-x)^{n-2} \quad \text{where } y \geq x$$

Problem 6

X_1, X_2, X_3 are three independent identically distributed continuous random variables with the following common pdf:

$$f(x) = \frac{8}{x^3} \quad \text{where } x \ge 2$$

Calculate the joint pdf for $X_{(1)}$ and $X_{(2)}$

Solution

$$F_X(x) = \int_2^x f_x(t)dt = \int_2^x \frac{8}{t^3}dt = 1 - \frac{4}{x^2}$$

$$f_{X_{(r)}, X_{(s)}}(x, y) =$$
$$\frac{n!}{(r-1)!(s-r-1)!(n-s)!}\left[F_X(x)\right]^{r-1}\left[F_X(y) - F_X(x)\right]^{s-r-1}\left[1 - F_X(y)\right]^{n-s} f_X(x) f_X(y)$$

The joint pdf for $X_{(1)}$ and $X_{(2)}$, $r=1$, $s=2$, $n=3$.

$$f_{X_{(1)}, X_{(2)}}(x, y)$$
$$= \frac{n!}{(r-1)!(s-r-1)!(n-s)!}\left[F_X(x)\right]^{r-1}\left[F_X(y) - F_X(x)\right]^{s-r-1}\left[1 - F_X(y)\right]^{n-s} f_X(x) f_X(y)$$
$$= \frac{3!}{(1-1)!(2-1-1)!(3-2)!}\left[F_X(x)\right]^{1-1}\left[F_X(y) - F_X(x)\right]^{2-1-1}\left[1 - F_X(y)\right]^{3-2} f_X(x) f_X(y)$$

$$= 6\left[1 - F_X(y)\right] f_X(x) f_X(y) \quad = 6\left(\frac{4}{y^2}\right)\left(\frac{8}{x^3}\right)\frac{8}{y^3} = \frac{1{,}536}{x^3 y^5} \quad \text{where } 2 \le x \le y$$

We can find $f_{X_{(1)}}(x)$ from $f_{X_{(1)}, X_{(2)}}(x, y)$.

$$f_{X_{(1)}}(x) = \int_x^{+\infty} f_{X_{(1)}, X_{(2)}}(x, y)dy = \int_x^{+\infty} \frac{1{,}536}{x^3 y^5}dy = \frac{1{,}536}{x^3}\int_x^{+\infty} y^{-5}dy = \frac{1{,}536}{x^3}\left[\frac{y^{-4}}{-5+1}\right]_x^{+\infty} = \frac{384}{x^7}$$

This result is OK; Problem 3 also gives us

$$f_{X_{(1)}}(x) = 3f_X(x)\left[1 - F_X(x)\right]^2 = 3\left(\frac{8}{x^3}\right)\left[1 - \left(1 - \frac{4}{x^2}\right)\right]^2 = \frac{384}{x^7} \text{ where } x \geq 2$$

We can find $f_{X_{(2)}}(y)$ from $f_{X_{(1)},X_{(2)}}(x, y)$.

$$f_{X_{(2)}}(y) = \int_2^y f_{X_{(1)},X_{(2)}}(x, y)\,dx = \int_2^y \frac{1,536}{x^3 y^5}\,dx = \frac{1,536}{y^5}\int_2^y x^{-3}dy = \frac{1,536}{y^5}\left[\frac{y^{-2}}{-2}\right]_2^y$$

$$= \frac{768}{y^5}\left(\frac{1}{4} - \frac{1}{y^2}\right) = \frac{192}{y^5}\left(1 - \frac{4}{y^2}\right)$$

The result is correct; from Problem 3, we also have:

$$f_{X_{(2)}}(x) = 6\left(\frac{8}{x^3}\right)\left(1 - \frac{4}{x^2}\right)\left(\frac{4}{x^2}\right) = \frac{192}{x^5}\left(1 - \frac{4}{x^2}\right)$$

The final topic on order statistics.

The joint pdf for $X_{(1)}$, $X_{(2)}$, $X_{(3)}$, …., $X_{(n)}$ is

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots, x_n) = \begin{cases} n!\,f_X(x_1)\,f_X(x_2)\ldots f_X(x_n) & \text{if } x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n \\ 0 & \text{if otherwise} \end{cases}$$

To illustrate the proof, I'll set $n = 3$ and prove that

$$f_{X_{(1)},X_{(2)},X_{(3)}}(x, y, z) = \begin{cases} 3!\,f_X(x)\,f_X(x)\,f_X(z) & \text{if } x \leq y \leq z \\ 0 & \text{if otherwise} \end{cases}$$

The proof for $n > 3$ is similar.

$$f_{X_{(1)},X_{(2)},X_{(3)}}(x,y,z)\,dx\,dy\,dz = P\left[x < X_{(1)} < x+dx, \quad y < X_{(2)} < y+dy, \quad z < X_{(1)} < z+dz\right]$$

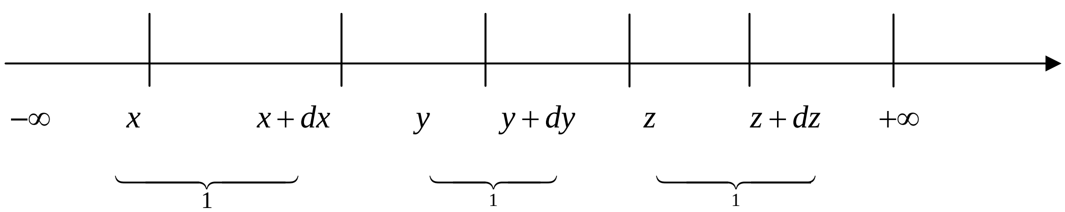
Once again, we can't write

$$f_{X_{(1)},X_{(2)},X_{(3)}}(x,y,z) = P\left[x < X_{(1)} < x+dx, \quad y < X_{(2)} < y+dy, \quad z < X_{(1)} < z+dz\right]$$

The above expression is wrong because $f_{X_{(1)},X_{(2)},X_{(3)}}(x,y,z)$ is not a real probability. However, $f_{X_{(1)},X_{(2)},X_{(3)}}(x,y,z)\,dx\,dy\,dz$ is a real probability.

Let's continue.

The only way to have $x < X_{(1)} < x+dx$, $y < X_{(2)} < y+dy$, and $z < X_{(3)} < z+dz$ where $x \le y \le z$ is to have one X in $(x, x+dx)$, another X in $(y, y+dy)$, and the last X in $(z, z+dz)$.



$$f_{X_{(1)},X_{(2)},X_{(3)}}(x,y,z)\,dx\,dy\,dz$$

$$= P\left[x < X_{(1)} < x+dx, \quad y < X_{(2)} < y+dy, \quad z < X_{(3)} < z+dz\right]$$

$$= \underbrace{\frac{3!}{1!1!1!}}_{\text{\# of permutations}} \quad \underbrace{f_X(x)\,dx}_{1\,X\,\in\,(x,\,x+dx)} \quad \underbrace{f_X(y)\,dy}_{1\,X\,\in\,(y,\,y+dy)} \quad \underbrace{f_X(z)\,dz}_{1\,X\,\in\,(z,\,z+dz)}$$

$$= 3!\, f_X(x)\, f_X(y)\, f_X(z)\,dx\,dy\,dz \quad \text{where } x \le y \le z$$

$$\Rightarrow \quad f_{X_{(1)},X_{(2)},X_{(3)}}(x,y,z) = 3!\, f_X(x)\, f_X(y)\, f_X(z)$$

Because $X_{(1)} \le X_{(2)} \le X_{(3)}$, then $f_{X_{(1)},X_{(2)},X_{(3)}}(x,y,z) = 0$ if $x \le y \le z$ is not satisfied.

Generally,

$$
\begin{aligned}
& f_{X_{(1)},\, X_{(2)},\dots,\, X_{(n)}}\left(x_1, x_2, \dots, x_n\right) dx_1 dx_2 \dots dx_n \\
&= P\left[x_1 < X_{(1)} < x_1 + dx_1, \quad x_2 < X_{(2)} < x_2 + dx_2,\ \dots,\ x_n < X_{(n)} < x_n + dx_n\right]
\end{aligned}
$$

To only way to have $x_1 < X_{(1)} < x_1 + dx_1, \quad x_2 < X_{(2)} < x_2 + dx_2,\ \dots,\ x_n < X_{(n)} < x_n + dx_n$ where $x_1 \le x_2 \le x_3 \le \dots \le x_n$ is to have one X in $\left(x_1, x_1 + dx_1\right)$, one X in $\left(x_2, x_2 + dx_2\right)$, …., one X in $\left(x_n, x_n + dx_n\right)$. The # of permutations is $n!$.

Consequently, for $x_1 \le x_2 \le x_3 \le \dots \le x_n$

$$
\begin{aligned}
& f_{X_{(1)},\, X_{(2)},\dots,\, X_{(n)}}\left(x_1, x_2, \dots, x_n\right) dx_1 dx_2 \dots dx_n \\
&= P\left[x_1 < X_{(1)} < x_1 + dx_1, \quad x_2 < X_{(2)} < x_2 + dx_2,\ \dots,\ x_n < X_{(n)} < x_n + dx_n\right] \\
&= n!\, f_X\left(x_1\right) dx_1\, f_X\left(x_2\right) dx_2 \dots f_X\left(x_n\right) dx_n
\end{aligned}
$$

$$
\Rightarrow \quad f_{X_{(1)},\, X_{(2)},\dots,\, X_{(n)}}\left(x_1, x_2, \dots, x_n\right) = n!\, f_X\left(x_1\right) f_X\left(x_2\right) \dots f_X\left(x_n\right)
$$

Problem 7

X_1, X_2, X_3 are three independent identically distributed continuous random variables with the following common pdf:

$$
f\left(x\right) = \frac{8}{x^3} \quad \text{where } x \ge 2
$$

Calculate the joint pdf for $X_{(1)}$, $X_{(2)}$, and $X_{(3)}$

Solution

If $x_1 \le x_2 \le x_3$, then

$$
f_{X_{(1)},\, X_{(2)}, X_{(3)}}\left(x_1, x_2, x_3\right) = 3!\, f_X\left(x_1\right) f_X\left(x_2\right) f_X\left(x_3\right) = 6\left(\frac{8}{x_1^3}\right)\left(\frac{8}{x_2^3}\right)\left(\frac{8}{x_3^3}\right)
$$

Otherwise,

$$f_{X_{(1)},X_{(2)},X_{(3)}}\left(x_1,x_2,x_3\right)=0$$

Problem 8

X_1 and X_2 are independent identically distributed exponential random variables with mean of 1. Find the pdf for $X_{(2)}-X_{(1)}$, the difference between the 2nd order statistics and the 1st order statistics.

Solution

X_1 and X_2 are independent identically distributed with the following common pdf $f_X\left(x\right)=e^{-x}$. The joint pdf for $X_{(1)}$ and $X_{(2)}$ is:

$$f_{X_{(1)},X_{(2)}}\left(x,y\right)=2!\,f_X\left(x\right)f_X\left(y\right)=2e^{-x}e^{-y}=2e^{-(x+y)} \quad \text{where } 0\le x\le y$$

We'll use the Jacobian method to find the joint pdf. Let $Y_1=X_{(1)}$, $Y_2=X_{(2)}-X_{(1)}$. Y_1 is just a fake random variable; we'll get rid of it in the end.

Recover the old random variables:

$$X_{(1)}=Y_1,\quad X_{(2)}=Y_1+Y_2$$

$$\Rightarrow\quad f_{X_{(1)},X_{(2)}}\left(x,y\right)=2e^{-(x+y)}=2e^{-(2y_1+y_2)}$$

$$\Rightarrow\quad \frac{\delta x_{(1)}}{\delta y_1}=1,\ \frac{\delta x_{(1)}}{\delta y_2}=0,\ \frac{\delta x_{(2)}}{\delta y_1}=1,\ \frac{\delta x_{(2)}}{\delta y_2}=1$$

$$\Rightarrow\quad J=\begin{vmatrix}\dfrac{\delta x_{(1)}}{\delta y_1} & \dfrac{\delta x_{(1)}}{\delta y_2}\\ \dfrac{\delta x_{(2)}}{\delta y_1} & \dfrac{\delta x_{(2)}}{\delta y_2}\end{vmatrix}=\begin{vmatrix}1 & 0\\ 1 & 1\end{vmatrix}=1,\qquad \left|J\right|=1$$

$$\Rightarrow\quad f_{Y_1,Y_2}\left(y_1,y_2\right)=\left|J\right|f_{X_{(1)},X_{(2)}}\left(x,y\right)=2e^{-(2y_1+y_2)} \quad \text{where } y_1\ge 0 \text{ and } y_2\ge 0.$$

Next, we get rid of Y_1:

$$f_{Y_2}(y_2) = \int f_{Y_1,Y_2}(y_1, y_2) dy_1 = \int_0^{+\infty} 2e^{-(2y_1+y_2)} dy_1 = 2e^{-y_2} \int_0^{+\infty} e^{-2y_1} dy_1 = e^{-y_2}$$

Incidentally, we notice that $f_{Y_1,Y_2}(y_1, y_2) = 2e^{-(2y_1+y_2)} = 2e^{-2y_1} e^{-y_2} = g(y_1) h(y_2)$, where $g(y_1) = 2e^{-2y_1}$ and $h(y_2) = e^{-y_2}$. So we know that Y_1 and Y_2 are independent.

Homework: Redo all the problems in this chapter.

Chapter 32 Double expectation

Formula $E(X) = E_Y\left[E_{X|Y}(X|Y)\right]$

Let's use a simple example to understand the meaning of the above formula. Let X =a group of college students' SAT scores, let Y =gender of a college student (either male or female).

Then the above formula becomes:

$$\begin{aligned}&E(\text{SAT scores})\\&=E_{GENDER}\left[E_{\text{SAT Scores}|GENDER}(\text{SAT Scores}|GENDER)\right]\end{aligned}$$

The formula says that to find the average of a group of college students' SAT scores, we first find

$$E_{\text{SAT Scores}|GENDER}(\text{SAT Scores}|GENDER).$$

The above symbol means the average SAT score is conditioned on gender. In other words, we are dividing the college students into two groups by gender --- male college students and female college students and finding the average SAT score of male students and of female students.

Next, we find

$$E_{GENDER}\left[E_{\text{SAT Scores}|GENDER}(\text{SAT Scores}|GENDER)\right]$$

The above symbol means that we are calculating the weighted average of the male students' average SAT score and the female students' average SAT score. Intuitively, the weighted average for these two averages should be the overall average SAT score of males and females combined, as expressed in the formula below:

$$\begin{aligned}&E(\text{SAT scores})\\&=E_{GENDER}\left[E_{\text{SAT Scores}|GENDER}(\text{SAT Scores}|GENDER)\right]\\&=\Pr(\text{male})\times E(\text{SAT Scores}|\text{GENDER=male})\\&\quad+\Pr(\text{female})\times E(\text{SAT Scores}|\text{GENDER=female})\end{aligned}$$

Of course, we can divide the college students by other categories. For example, we can divide them by major (English, Philosophy, Math, …). We then calculate the averaged SAT score by major. The weighted average SAT score by each major should be the overall average SAT score for the college students as a whole.

$$\begin{aligned}
& E(\text{SAT scores}) \\
&= E_{\text{MAJOR}}\left[E_{\text{SAT Scores}\mid\text{MAJOR}}(\text{SAT Scores}\mid\text{MAJOR})\right] \\
&= \quad \Pr(\text{English Major})\times E(\text{SAT Scores}\mid\text{Major=English}) \\
&\quad + \Pr(\text{History Major})\times E(\text{SAT Scores}\mid\text{Major=History}) \\
&\quad + \dots
\end{aligned}$$

Problem 1

A group of 20 graduate students (12 male and 8 female) have a total GRE score of 12,940. The GRE score distribution by gender is as follows:

Total GRE scores of 12 males	7,740
Total GRE scores of 8 females	5,200
Total GRE score	12,940

Find the average GRE score twice. For the first time, do not use the double expectation theorem. The second time, use the double expectation theorem. Show that you get the same result.

Solution

(1) Find the mean without using the double expectation theorem.

Average GRE score for 20 graduate students

$$=\frac{\text{Total GRE scores}}{\#\text{ of students}}=\frac{12,940}{20}=647$$

(2) Find the mean using the double expectation theorem.

$$\begin{aligned}
& E(\text{GRE scores}) \\
&= E_{GENDER}\left[E_{\text{GRE Scores}\mid GENDER}(\text{GRE Scores}\mid GENDER)\right] \\
&= \quad \Pr(\text{male})\times E(\text{GRE Scores}\mid\text{GENDER=male}) \\
&\quad + \Pr(\text{female})\times E(\text{GRE Scores}\mid\text{GENDER=female})
\end{aligned}$$

Pr(male)=12/20=0.6, Pr(female)=8/20=0.4

$\text{E(GRE Scores}\mid\text{GENDER=male)}$=7,740/12=645

$E(\text{GRE Scores} | \text{GENDER=female}) = 5{,}200/8=650$

$E(\text{GRE Scores})=0.6(645)+0.4(650)=647$

You can see the two methods produce an identical result.

Problem 2 (This problem is one of the more difficult problems. If you can calculate $E(N)$, most likely that is good enough for Exam P.)

The number of claims, N, incurred by a policyholder has the following distribution:

$$\Pr(N = n) = C_3^n p^n (1-p)^{3-n} \qquad n = 0, 1, 2, 3$$

p is uniformly distributed over [0, 1]. Find $E(N), Var(N)$.

Solution

If p is constant, N has the binomial distribution with mean and variance:

$$E(N) = 3p, \quad Var(N) = 3p(1-p) \text{ -- if } p \text{ is constant}$$

However, p is not constant. So we cannot directly use the above formula. What should we do? In situations like this, the double expectation theorem comes in handy.

To find $E(N)$, we divide N into different groups by p --- just as we divide the college students into male and female students, except this time we have an infinite number of groups (P is a continuous random variable). Each value of $P = p$ is a separate group. For each group, we will calculate its mean. Then we will find the weighted average mean for all the groups, with weight being the probability of each group's *p* value. The result should be $E(N)$.

$$E(N) = E_P\left[E_{N|P}(N | P = p)\right]$$

$E_{N|P}(N | P = p) = 3p$ (This the mean for a given group with $P = p$)

Next, we need to find the weighted average of each group's mean:

$$E(N) = \int_0^1 \underbrace{1}_{\substack{\text{Probability of} \\ \text{each group's} \\ \text{occurrence}}} \times \underbrace{3p}_{\substack{\text{each group's} \\ \text{mean for N}}} \, dp = \left[\frac{3}{2}p^2\right]_0^1 = \frac{3}{2}$$

The integration is needed because we have an infinite number of groups.

Alternatively,

$$E(N) = E_P\left[E_{N \mid P}(N \mid P = p)\right] = E_P(3p) = 3E_P(p) = 3 \times \frac{1}{2} = \frac{3}{2}$$

In the above, $E_P(p) = \frac{1}{2}$ because P is uniform over [0, 1]

Next, we find $Var(N)$ using the formula $Var(N) = E(N^2) - E^2(N)$.

We can calculate $E(N^2)$ the same way we calculated $E(N)$.

$$E(N^2) = E_P\left[E_{N \mid P}(N^2 \mid P = p)\right]$$

For a given group $P = p$, N is binomial with a mean of $3p$ and a variance of $3p(1-p)$.

$$E_{N \mid P}(N^2 \mid P = p) = E^2_{N \mid P}(N \mid P = p) + Var_{N \mid P}(N \mid P = p)$$
$$= (3p)^2 + 3p(1-p) = 6p^2 + 3p$$

Next, we calculate the weighted average of $E_{N \mid P}(N^2 \mid P = p)$ of all the groups, with weight being the probability of each group:

$$E(N^2) = \int_0^1 \underbrace{1}_{\substack{\text{Probability of} \\ \text{each group's} \\ \text{occurrence}}} \times \underbrace{6p^2 + 3p}_{\substack{\text{each group's} \\ \text{mean for } N^2}} \, dp = \left[2p^3 + \frac{3}{2}p^2\right]_0^1 = 2 + \frac{3}{2}$$

$$Var(N) = E(N^2) - E^2(N) = 2 + \frac{3}{2} - \left(\frac{3}{2}\right)^2 = \frac{5}{4}$$

Alternatively, you can use the following formula (if you have memorized it):

$$Var(X) = E_Y\left[Var_{X|Y}(X|Y=y)\right] + Var_Y\left[E_{X|Y}(X|Y=y)\right]$$

To help memorize this formula, we can rewrite it as

$$Var(X) = EV + VE \text{, where } EV = E_Y\left[Var(X|Y)\right],\ VE = Var_Y\left[E(X|Y)\right]$$

If you have extra brainpower, you can learn the above formula. However, if you need to work on more basic concepts and problems, forget about this formula – you are better off learning something else.

Applying the variance formula, we have:

$$Var(N) = E_P\left[Var_{N|P}(N|P=p)\right] + Var_P\left[E_{N|P}(N|P=p)\right]$$

Because $N|P=p$ is binomial with parameter 3 and p, we have

$$E_{N|P}(N|P) = 3p,\ Var_{N|P}(N|P=p) = 3p(1-p)$$

$$E_P\left[Var_{N|P}(N|P=p)\right] = E_P\left[3p(1-p)\right] = 3E_P\left[p-p^2\right] = 3\left[E_P(p) - E_P(p^2)\right]$$
$$Var_P\left[E_{N|P}(N|P=p)\right] = Var_P\left[3p\right] = 9Var_P\left[p\right]$$

Because P is uniform over [0, 1],

$$E_P(p) = \frac{1}{2},\quad Var_P\left[p\right] = \left(\frac{1}{2\sqrt{3}}\right)^2 = \frac{1}{12},\quad E_P(p^2) = \left(\frac{1}{2}\right)^2 + \frac{1}{12}$$

You should remember that if X is uniform over $[a,b]$, then

$$E(X) = \frac{a+b}{2},\quad \sigma_X = \frac{b-a}{2\sqrt{3}}$$

$$3\left[E_P(p) - E_P(p^2)\right] = 3\left[\frac{1}{2} - \left(\frac{1}{2}\right)^2 - \frac{1}{12}\right] = \frac{1}{2}$$

$$9Var_P\left[p\right] = 9 \times \frac{1}{12} = \frac{3}{4}$$

$$Var(N) = E_P\left[Var_{N|P}(N|P=p)\right] + Var_P\left[E_{N|P}(N|P=p)\right] = \frac{1}{2} + \frac{3}{4} = \frac{5}{4}$$

Problem 3

X is a Poisson random variable with mean λ. λ is uniformly distributed over [0, 6]. Calculate $E(X)$ and $Var(X)$.

Solution

$X|\lambda$ is Poisson with parameter λ. So $E(X|\lambda) = Var(X|\lambda) = \lambda$.

$$E(X) = E_\lambda\left[E(X|\lambda)\right] = E_\lambda[\lambda] = \frac{6}{2} = 3$$

$$Var(X) = Var_\lambda\left[E(X|\lambda)\right] = Var_\lambda[\lambda] = \frac{(6-0)^2}{12} = 3$$

Homework for you: #20, May 2000; #10, May2001.

Chapter 33 Moment generating function

Moment generating function (MGF) is one of the least intuitive concepts in probability theories. Before we bother to memorize a bunch of MGF formulas, let's understand why MGF was invented and what's the use of it.

MGF is to probability theories as catalysts are to chemical reactions. If catalysts are not used in a chemical reaction, the chemical reaction can still take place but it may take a long time for it to occur. However, if catalysts are used, a chemical reaction can proceed quickly. Catalysts reduce the amount of energy needed to start a chemical reaction.

MGF reduces the amount of energy we need to have to get things done in probability theories.

Example 1. The # of accidents that happen on a particular highway during a 3-month period (i.e. quarter) is a Poisson random variable.

On average,

- 1 accident happens during the 1st quarter
- 2 accidents happen during the 2nd quarter
- 3 accidents happen during the 3rd quarter
- 4 accidents happen during the 4th quarter

Assume that the # of accidents in any quarter is independent of the # of accidents in any other quarter. Calculate the probability that at least 4 accidents happen on the highway in a year.

Solution

First, let's define 5 random variables:

- N_1 is the # of accident in the 1st quarter. N_1 is Poisson with $\lambda_1 = 1$.
- N_2 is the # of accident in the 2nd quarter. N_2 is Poisson with $\lambda_2 = 2$.
- N_3 is the # of accident in the 3rd quarter. N_3 is Poisson with $\lambda_3 = 3$.
- N_4 is the # of accident in the 4th quarter. N_4 is Poisson with $\lambda_4 = 4$.
- N is the # of accident in a year. $N = N_1 + N_2 + N_3 + N_4$

We are asked to find $P(N \geq 4)$. To find $P(N \geq 4)$, we'll need to find the probability density function of N. We can guess that N is a Poisson random variable with parameter $\lambda = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 10$. But how can we prove that N is a Poisson random variable with $\lambda = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4$? **Proving this will be difficult with out MGF. However, the proof will be a piece of cake if we use MGF, as you'll soon see.**

Example 2. Random variables X and Y are two independent normal random variables. We want to know whether their sum, $X+Y$, is also a normal random variable. How can we quickly check whether $X+Y$ is also normal?

With MGF, we can prove, effortlessly, that $X+Y$ is also normal. Without MGF, the proof will take lot of work.

Key point:

MGF enables us to quickly find the distribution of the sum of n independent random variables.

This point will be made clear to you later. For now, let's define MGF and write up some key formulas.

14 Key MGF formulas you must memorize

1. $M_X(t)=E\left(e^{tX}\right)$**.** This is the definition of MFG.

If X is discrete, then $M_X(t)=E\left(e^{tX}\right)=\sum_x e^{tx}p_X(x)$

If X is continuous, then $M_X(t)=E\left(e^{tX}\right)=\int_{-\infty}^{+\infty} e^{tx}f_X(x)dx$

2. $M_{aX}(t)=E\left[e^{(aX)t}\right]=E\left[e^{X(at)}\right]=M_X(at)$

3. $M_b(t)=E\left[e^{bt}\right]=e^{bt}$. Here you can think of b as being a random variable that takes on value b only.

4. If X and Y are independent, then $M_{X+Y}(t)=M_X(t)M_Y(t)$

The proof is simple. $M_{X+Y}(t)=E\left[e^{(X+Y)t}\right]=E\left[e^{Xt}e^{Yt}\right]$

Since X and Y are independent, $E\left[e^{Xt}e^{Yt}\right]=E\left[e^{Xt}\right]E\left[e^{Yt}\right]=M_X(t)M_Y(t)$

Generally, if $Y=X_1+X_2+...+X_n$ where X_1, $X_2,\ldots,X_n$ are independent, then

$$M_Y(t)=M_{X_1}(t)M_{X_2}(t)...M_{X_n}(t)$$

$$M_{aX+b}(t)=E\left[e^{t(aX+b)}\right]=E\left[e^{t(aX)}e^{tb}\right]=e^{tb}E\left[e^{t(aX)}\right]=e^{tb}M_{aX}(t)$$

Alternatively, imagine you have two independent random variables aX and b.

Then $M_{aX+b}(t) = M_{aX}(t) M_b(t) = M_{aX}(t) e^{tb} = e^{tb} M_{aX}(t)$

5. $M_{\frac{X+b}{a}}(t) = M_{\frac{X}{a}+\frac{b}{a}}(t) = M_{\left(\frac{1}{a}\right)X}(t) M_{\frac{b}{a}}(t) = M_X\left(\frac{t}{a}\right) e^{\frac{b}{a}t} = e^{\frac{b}{a}t} M_X\left(\frac{t}{a}\right)$

6. $\left[M_X(t)\right]_{t=0} = 1$. To see why, notice that $\left[M_X(t)\right]_{t=0} = E\left(e^{0*X}\right) = E\left(e^0\right) = E(1) = 1$.

7. $\left[\frac{d}{dt} M_X(t)\right]_{t=0} = E(X)$, $\left[\frac{d^2}{dt^2} M_X(t)\right]_{t=0} = E\left(X^2\right)$, $\left[\frac{d^n}{dt^n} M_X(t)\right]_{t=0} = E\left(X^n\right)$

To see why, please note

$$e^{tX} = 1 + \frac{1}{1!}(tX) + \frac{1}{2!}(tX)^2 + \frac{1}{3!}(tX)^3 + ... + \frac{1}{n!}(tX)^n + ... \quad \text{(Taylor series)}$$

Taking expectation regarding to X from both sides:

$$M_X(t) = E\left(e^{tX}\right) = 1 + tE(X) + \frac{1}{2!}t^2 E\left(X^2\right) + \frac{1}{3!}t^3\left(X^3\right) + ... + \frac{1}{n!}t^n\left(X^n\right) + ...$$

Then we have:

$$\left[\frac{d}{dt} M_X(t)\right]_{t=0} = E(X), \quad \left[\frac{d^2}{dt^2} M_X(t)\right]_{t=0} = E\left(X^2\right), \quad ..., \quad \left[\frac{d^n}{dt^n} M_X(t)\right]_{t=0} = E\left(X^n\right)$$

8. MGF for Bernoulli distribution

This is a special case of binominal distribution with $n = 1$ (i.e. the # of trial is one).

$$X = \begin{cases} 1 & \text{with probability of } p \\ 0 & \text{with probability of } q = 1 - p \end{cases}$$

MGF	$M_X(t) = E\left(e^{tX}\right) = pe^{t(1)} + qe^{t(0)} = pe^t + q$

9. MGF for binomial distribution

Probability mass function $\qquad p_X(x) = C_n^x p^x q^{n-x}$

MGF	$M_X(t) = E\left(e^{tX}\right) = \left(pe^t + q\right)^n$

How to memorize MGF

Binomial distribution is the sum of n independent identically distributed Bernoulli random variables. Let X be the binomial random variable with parameter n and p. Let Y_1, Y_2,..., Y_n be n independent identically distributed Bernoulli random variables with parameter p. Then

$$X = Y_1 + Y_2 + ... + Y_n$$

For example. Let X represent the total # of heads you get if you throw a coin 10 times. Then

Y_1 is the # of head you get in the 1st throw; Y_1 is either 0 or 1.
Y_2 is the # of heads you get in the 2nd throw; Y_2 is either 0 or 1.
……

Y_{10} is the # of heads you get in the 10th throw; Y_{10} is either 0 or 1.

Then clearly $X = Y_1 + Y_2 + ... + Y_{10}$.

$$M_X(t) = M_{Y_1+Y_2+...+Y_n}(t) = M_{Y_1}(t) M_{Y_2}(t) ... M_{Y_n}(t)$$

$$M_{Y_1}(t) = pe^t + q$$
$$M_{Y_2}(t) = pe^t + q$$
……
$$M_{Y_n}(t) = pe^t + q$$

So we have:

$$M_X(t) = M_{Y_1+Y_2+...+Y_n}(t) = M_{Y_1}(t) M_{Y_2}(t) ... M_{Y_n}(t) = \left(pe^t + q\right)^n$$

10. MGF for Poisson distribution

$$p_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \text{ where } x = 0, 1, 2, ...$$

$$M_X(t) = E\left(e^{tX}\right) = \sum_{x=0}^{+\infty} e^{tx}\left(e^{-\lambda}\frac{\lambda^x}{x!}\right) = e^{-\lambda}\sum_{x=0}^{+\infty}\frac{\left(\lambda e^t\right)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda\left(e^t - 1\right)}$$

$$M_X(t) = e^{\lambda\left(e^t - 1\right)}$$

11. MGF for a standard normal distribution

If X is the standard normal distribution with mean of 0 and variance of 1, then

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},$$

$$M_X(t) = E\left[e^{tX}\right] = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2 + tx} dx$$

$$-\frac{1}{2}x^2 + tx = -\frac{1}{2}\left(x^2 - 2tx\right) = -\frac{1}{2}\left[\left(x^2 - 2tx + t^2\right) - t^2\right] = -\frac{1}{2}\left[(x-t)^2 - t^2\right]$$

$$\Rightarrow \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2 + tx} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left[(x-t)^2 - t^2\right]} dx = e^{\frac{1}{2}t^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left[(x-t)^2\right]} dx$$

Setting $Y = X - t$, we have: $\displaystyle\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left[(x-t)^2\right]} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$

Because $\dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$ is the density function of a standard normal distribution, we have:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = 1$$

$$\Rightarrow \quad M_X(t) = E\left[e^{tX}\right] = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = e^{\frac{1}{2}t^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left[(x-t)^2\right]} dx = e^{\frac{1}{2}t^2}$$

If X is a standard normal random variable, then $M_X(t) = e^{\frac{1}{2}t^2}$

12. MGF for a normal random variable with mean μ and standard deviation σ

Let X represent a normal random variable with mean μ and standard deviation σ. Let Z represent the standard normal random variable. Then we have:

$$Z=\frac{X-\mu}{\sigma} \quad \Rightarrow \quad X=\mu+\sigma Z$$

$$\boxed{M_X(t)=M_{\mu+\sigma Z}(t)=e^{\mu t}M_{\sigma Z}(t)=e^{\mu t}M_Z(\sigma t)=e^{\mu t+\frac{1}{2}\sigma^2 t^2}}$$

13. MGF for Exponential distribution

If X is exponentially distributed with parameter λ, then

$f_X(x)=\lambda e^{-\lambda x}$, where $x \geq 0$

$$\boxed{M_X(t)=E\left(e^{tX}\right)=\int_0^{+\infty} e^{tx}\lambda e^{-\lambda x}dx=\int_0^{+\infty}\lambda e^{-(\lambda-t)x}dx=\frac{\lambda}{\lambda-t}=\frac{1}{1-t/\lambda}=\frac{1}{1-\theta t}}$$

14. MGF for gamma distribution

Let X_1, X_2,..., X_n represent n independent identically distributed exponential random variable with a common parameter λ. Let $S=\sum_{i=1}^{n} X_i$. As explained before, Y has gamma distribution with the following pdf:

$$f_Y(y)=\lambda\frac{(\lambda y)^n}{(n-1)!}e^{-(\lambda y)}, \quad \text{where } y \geq 0$$

The MGF of Y is:

$$\boxed{\begin{array}{l} M_Y(t)=M_{X_1+X_2+...+X_n}(t)=M_{X_1}(t)M_{X_2}(t)...M_{X_n}(t)=\left(\dfrac{\lambda}{\lambda-t}\right)^n=\left(\dfrac{1}{1-\theta t}\right)^n, \\ \text{where } t<\lambda=\dfrac{1}{\theta} \end{array}}$$

MGF's for geometric distribution, negative binomial, and uniform are more complex and less frequently tested. I recommend that you don't memorize their moment generating functions. However, you might want to memorize the MFG for Bernoulli, binominal, exponential, gamma, Poisson, and normal. **At minimum, memorize the MGF for exponential and Poisson. Exponential and Poisson distributions have elegant MGF's and are frequently tested in Exam P.**

Sample Problems and Solutions

Problem 1

Random variable X has the following distribution:

$X = x$	$p(x)$
0	0.1
1	0.4
2	0.5

Find $M_X(t)$.

Solution

This problem simply tests your knowledge of the definition of the moment generating function.

$$M_X(t) = E(e^{tX}) = \sum p(x)e^{tx} = 0.1e^{t0} + 0.4e^{t1} + 0.5e^{t2} = 0.1 + 0.4e^t + 0.5e^{2t}$$

Problem 2

A random variable has the following moment generating function:

$$M_X(t) = \frac{1}{7}e^t + \frac{2}{7}e^{2t} + \frac{4}{7}e^{3t}$$

Find the pdf, mean and variance of X.

Solution

Because $M_X(t) = E(e^{tX})$, we have

$$f(x) = \begin{cases} \frac{1}{7} & x = 1 \\ \frac{2}{7} & x = 2 \\ \frac{4}{7} & x = 3 \end{cases}$$

$$E(X) = \left[\frac{d}{dt}M_X(t)\right]_{t=0} = \left\{\frac{d}{dt}\left[\frac{1}{7}e^t + \frac{2}{7}e^{2t} + \frac{4}{7}e^{3t}\right]\right\}_{t=0}$$

$$=\left[\frac{1}{7}(1e^{t})+\frac{2}{7}(2e^{2t})+\frac{4}{7}(3e^{3t})\right]_{t=0}=\frac{1}{7}(1)+\frac{2}{7}(2)+\frac{4}{7}(3)=\frac{17}{7}$$

$$E(X^2)=\left[\frac{d^2}{dt^2}M_X(t)\right]_{t=0}=\left\{\frac{d}{dt}\left[\frac{d}{dt}M_X(t)\right]\right\}_{t=0}$$

$$=\left\{\frac{d}{dt}\left[\frac{1}{7}(1e^{t})+\frac{2}{7}(2e^{2t})+\frac{4}{7}(3e^{3t})\right]\right\}_{t=0}=\left[\frac{1}{7}(1^2e^{t})+\frac{2}{7}(2^2e^{2t})+\frac{4}{7}(3^3e^{3t})\right]_{t=0}$$

$$=\left[\frac{1}{7}(1^2)+\frac{2}{7}(2^2)+\frac{4}{7}(3^3)\right]=\frac{45}{7}$$

$$Var(X)=E(X^2)-E^2(X)=\frac{45}{7}-\left(\frac{17}{7}\right)^2=\frac{26}{49}$$

Problem 3

Prove that the linear combination of two independent normal random variables is also normal.

Solution

Assume

- X_1 is normal with mean of μ_1 and standard deviation of σ_1
- X_2 is normal with mean of μ_2 and standard deviation of σ_2

We need to prove that aX_1+bX_2 is also normal for $a\neq 0$.

$$M_{X_1}(t)=e^{\mu_1 t+\frac{1}{2}\sigma_1^2t^2},\quad M_{X_2}(t)=e^{\mu_2 t+\frac{1}{2}\sigma_2^2t^2}$$
$$\Rightarrow\quad M_{aX_1+bX_2}(t)=M_{aX_1}(t)M_{bX_2}(t)$$

$$M_{aX_1}(t)=M_{X_1}(at)=\exp\left[\mu_1(at)+\frac{1}{2}\sigma_1^2(at)^2\right]$$

$$M_{bX_2}(t)=M_{X_2}(bt)=\exp\left[\mu_2(bt)+\frac{1}{2}\sigma_2^2(bt)^2\right]$$

$$\Rightarrow M_{aX_1+bX_2}(t) = M_{aX_1}(t)M_{bX_2}(t) = \exp\left[\mu_1(at)+\frac{1}{2}\sigma_1^2(at)^2\right]\exp\left[\mu_2(bt)+\frac{1}{2}\sigma_2^2(bt)^2\right]$$

$$= \exp\left[\mu_1(at)+\frac{1}{2}\sigma_1^2(at)^2+\mu_2(bt)+\frac{1}{2}\sigma_2^2(bt)^2\right]$$

$$= \exp\left[(a\mu_1+b\mu_2)t+\frac{1}{2}(a^2\sigma_1^2+b^2\sigma_2^2)t^2\right]$$

$\exp\left[(a\mu_1+b\mu_2)t+\frac{1}{2}(a^2\sigma_1^2+b^2\sigma_2^2)t^2\right]$ is the MGF for a normal random variable that has the following mean and standard deviation:

The mean $\mu = a\mu_1 + b\mu_2$
The standard deviation $\sigma = \sqrt{a^2\sigma_1^2+b^2\sigma_2^2}$

You see aX_1+bX_2 is normal with $\mu = a\mu_1+b\mu_2$ and $\sigma = \sqrt{a^2\sigma_1^2+b^2\sigma_2^2}$.

Problem 4

Prove that the sum of two independent Poisson random variable is also a Poisson random variable.

Solution

Let N_1 and N_2 represent two independent Poisson random variable with λ_1 and λ_2 respectively. $M_{N_1}(t) = \exp\left[\lambda_1(e^t-1)\right]$. $M_{N_2}(t) = \exp\left[\lambda_2(e^t-1)\right]$

Because N_1 and N_2 are independent, we have:

$$M_{N_1+N_2}(t) = M_{N_1}(t)M_{N_2}(t) = \exp\left[\lambda_1(e^t-1)\right]\exp\left[\lambda_2(e^t-1)\right] = \exp\left[(\lambda_1+\lambda_2)(e^t-1)\right]$$

$\exp\left[(\lambda_1+\lambda_2)(e^t-1)\right]$ is the MGF for a Poisson random variable with parameter $\lambda_1+\lambda_2$.

If N_i is a Poisson random variable with mean λ_i (where $i = 1, 2, ..., k$), then $\sum_{i=1}^{k} N_i$ is also a Poisson random variable with mean $\sum_{i=1}^{k}\lambda_i$

Problem 5

The # of accidents that happen on a particular highway during a 3-month period (i.e. quarter) is a Poisson random variable.

On average,

- 1 accident happens during the 1st quarter
- 2 accidents happen during the 2nd quarter
- 3 accidents happen during the 3rd quarter
- 4 accidents happen during the 4th quarter

Assume that the # of accidents in any quarter is independent of the # of accidents in any other quarter. Calculate the probability that at least 4 accidents happen on the highway in a year.

Solution

First, let's define 5 random variables:

- N_1 is the # of accident in the 1st quarter. N_1 is Poisson with $\lambda_1 = 1$.
- N_2 is the # of accident in the 2nd quarter. N_2 is Poisson with $\lambda_2 = 2$.
- N_3 is the # of accident in the 3rd quarter. N_3 is Poisson with $\lambda_3 = 3$.
- N_4 is the # of accident in the 4th quarter. N_4 is Poisson with $\lambda_4 = 4$.
- N is the # of accident in a year. $N = N_1 + N_2 + N_3 + N_4$

$N = N_1 + N_2 + N_3 + N_4$ is also a Poisson random variable with mean $\lambda = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 10$.

$$P(N \geq 4) = 1 - \left[P(N=0) + P(N=1) + P(N=2) + P(N=3) \right]$$

$$= 1 - e^{-10} \left[\frac{10^0}{0!} + \frac{10^1}{1!} + \frac{10^2}{2!} + \frac{10^3}{3!} \right] = 0.99$$

Problem 6

$f_X(x) = \frac{1}{2} e^{-|x|}$, where $-\infty < x < +\infty$.

Calculate $M_X(t)$.

Solution

$$M_X(t)=E\left[e^{tX}\right]=\int_{-\infty}^{+\infty}e^{tx}f(x)dx=\int_{-\infty}^{+\infty}e^{tx}\frac{1}{2}e^{-|x|}dx=\frac{1}{2}\left[\int_{-\infty}^{0}e^{tx}e^{x}dx+\int_{0}^{+\infty}e^{tx}e^{-x}dx\right]$$

$$=\frac{1}{2}\left[\int_{-\infty}^{0}e^{(t+1)x}dx+\int_{0}^{+\infty}e^{(t-1)x}dx\right]=\frac{1}{2}\left[\frac{1}{t+1}e^{(t+1)x}\right]_{-\infty}^{0}+\frac{1}{2}\left[\frac{1}{t-1}e^{(t-1)x}\right]_{0}^{+\infty}$$

If $t+1>0$ or $t>-1$, $\left[\frac{1}{t+1}e^{(t+1)x}\right]_{-\infty}^{0}=\frac{1}{t+1}\left[e^{0}-e^{(t+1)(-\infty)}\right]=\frac{1}{t+1}[1-0]=\frac{1}{t+1}$.

If $t-1<0$ or $t<1$, $\left[\frac{1}{t-1}e^{(t-1)x}\right]_{0}^{+\infty}=\frac{1}{t-1}\left[e^{(t-1)\infty}-1\right]=\frac{1}{t-1}[0-1]=-\frac{1}{t-1}$

So for $-1<t<1$, we have:

$$M_X(t)=\frac{1}{2}\left[\frac{1}{t+1}e^{(t+1)x}\right]_{-\infty}^{0}+\frac{1}{2}\left[\frac{1}{t-1}e^{(t-1)x}\right]_{0}^{+\infty}=\frac{1}{2}\left(\frac{1}{t+1}-\frac{1}{t-1}\right)=\frac{1}{2}\left(\frac{1}{1+t}+\frac{1}{1-t}\right)=\frac{1}{1-t^2}$$

Problem 7

A discrete random variable X takes 3 possible values: 0, 1, and 2. The 1st moment of X is 1; the 2nd moment is 1.5.

Find $M_X(t)$

Solution

The 1st moment of X refers to $E(X)$. The 2nd moment of X refers to $E(X^2)$.
Generally, the k-th moment of X refers to $E(X^k)$.

Let a, b, and c represent the probability that X is equal to 0, 1, and 2 respectively.

$X=x$	0	1	2
$p_X(x)$	a	b	c

$$E(X)=0(a)+1(b)+2(c)=b+2c=1$$
$$E(X^2)=0^2(a)+1^2(b)+2^2(c)=b+4c=1.5$$

Solving the above equations, we have:

$$b=\frac{1}{2},\ c=\frac{1}{4},\ a=1-(b+c)=\frac{1}{4}$$

$$\Rightarrow \quad M_X(t)=ae^{0t}+be^{t}+ce^{2t}=\frac{1}{4}+\frac{1}{2}e^{t}+\frac{1}{4}e^{2t}$$

Problem 8

X is an exponential random variable with mean $\lambda=5$. $Y=2X+3$. Find $M_Y(t)$.

Solution

$$M_X(t)=\frac{\lambda}{\lambda-t}=\frac{5}{5-t}$$

$$M_Y(t)=M_{2X+3}(t)=M_{2X}(t)M_3(t)$$
$$M_{2X}(t)=M_X(2t)=\frac{5}{5-2t},\quad M_3(t)=e^{3t}$$
$$\Rightarrow\ M_Y(t)=M_{2X+3}(t)=M_{2X}(t)M_3(t)=\frac{5}{5-2t}e^{3t}.$$

Problem 9

$X=U+V$. U is exponentially distributed with parameter $\lambda=2$. V is a Poisson random variable with mean 3. U and V are independent.

Find $M_X(t)$.

Solution

$$M_U(t)=\frac{2}{2-t},\quad M_V(t)=\exp\left[3\left(e^{t}-1\right)\right]$$
$$M_X(t)=M_{U+V}(t)=M_U(t)M_V(t)=\frac{2}{2-t}\exp\left[3\left(e^{t}-1\right)\right]$$

Problem 10

$$M_Y(t) = \frac{5}{5-2t}e^{10t}$$

Find $E(Y)$ and $Var(Y)$

Solution

Method 1

$M_Y(t)$ is the product of two terms: $\dfrac{5}{5-2t}$ and e^{10t}.
$\dfrac{5}{5-2t}$ is the MGF for $2X$, where X is an exponential random variable with parameter $\lambda = 5$. e^{10t} is the MGF for 10. So $Y = 2X + 10$.

$$E(Y) = E(2X+10) = 2E(X) + 10 = 2\left(\frac{1}{5}\right) + 10 = 10.4$$

$$Var(Y) = Var(2X+10) = 4Var(X) = 4\left(\frac{1}{5}\right)^2 = 0.16$$

Please note that the mean and the standard deviation of an exponential distribution with parameter λ are both $\dfrac{1}{\lambda}$. In this problem, $E(X) = \sigma_X = \dfrac{1}{5}$

Method 2

$$M_Y(t) = \frac{5}{5-2t}e^{10t}$$

$$\frac{d}{dt}M_Y(t) = \frac{d}{dt}\left(\frac{5}{5-2t}e^{10t}\right) = \frac{5}{5-2t}\frac{d}{dt}e^{10t} + e^{10t}\frac{d}{dt}\frac{5}{5-2t}$$

$$= 10\left(\frac{5}{5-2t}\right)e^{10t} + e^{10t}5(2)(5-2t)^{-2} = \left(10 + \frac{2}{5-2t}\right)\left(\frac{5}{5-2t}\right)e^{10t}$$

$$= \left(10 + \frac{2}{5-2t}\right)M_X(t)$$

$$\Rightarrow \quad \frac{d}{dt}M_Y(t) = \left(10 + \frac{2}{5-2t}\right)M_X(t)$$

$$E(X) = \left[\frac{d}{dt}M_Y(t)\right]_{t=0} = \left[\left(10 + \frac{2}{5-2t}\right)M_X(t)\right]_{t=0} = \left(10 + \frac{2}{5-2t}\right)_{t=0}\left[M_X(t)\right]_{t=0} = 10.4$$

Please note that $\left[M_X(t)\right]_{t=0} = 1$

$$\frac{d^2}{dt^2}M_Y(t) = \frac{d}{dt}\left[\left(10 + \frac{2}{5-2t}\right)M_X(t)\right] = \left(10 + \frac{2}{5-2t}\right)\frac{d}{dt}M_X(t) + M_X(t)\frac{d}{dt}\left(10 + \frac{2}{5-2t}\right)$$

$$= \left(10 + \frac{2}{5-2t}\right)\frac{d}{dt}M_X(t) + M_X(t)4(5-2t)^{-2}$$

$$E(X^2) = \left[\frac{d^2}{dt^2}M_Y(t)\right]_{t=0} = \left[\left(10 + \frac{2}{5-2t}\right)\frac{d}{dt}M_X(t) + M_X(t)4(5-2t)^{-2}\right]_{t=0}$$

$$= \left(10 + \frac{2}{5-2t}\right)_{t=0}\left[\frac{d}{dt}M_X(t)\right]_{t=0} + \left[M_X(t)\right]_{t=0}\left[4(5-2t)^{-2}\right]_{t=0}$$

$$= \left(10 + \frac{2}{5}\right)10.4 + \frac{4}{5^2} = 10.4^2 + 0.16$$

$$Var(X) = E(X^2) - E^2(X) = 10.4^2 + 0.16 - 10.4^2 = 0.16$$

Problem 10

Random variable X has the following distribution:

$$X = \begin{cases} U & \text{with probability of 0.25} \\ V & \text{with probability of 0.75} \end{cases}$$

where

U is exponential distribution with parameter $\lambda = 1$
V is exponential distribution with parameter $\lambda = 2$

Find $M_X(t)$

Solution

$$M_X(t) = P(X = U)M_U(t) + P(X = V)M_V(t) = \frac{1}{4}\left(\frac{1}{1-t}\right) + \frac{3}{4}\left(\frac{2}{2-t}\right)$$

If the above solution seems difficult to understand, here is another solution:

$$P(X \le x) = P(X = U)P(U \le x) + P(X = V)P(V \le x) = \frac{1}{4}P(U \le x) + \frac{3}{4}P(V \le x)$$

Taking derivative regarding to x:

$$f(x) = \frac{d}{dx}P(X \le x) = \frac{d}{dx}\left[\frac{1}{4}P(U \le x) + \frac{3}{4}P(V \le x)\right] = \frac{1}{4}\frac{d}{dx}P(U \le x) + \frac{3}{4}\frac{d}{dx}P(V \le x)$$

$$= \frac{1}{4}f_U(x) + \frac{3}{4}f_V(x) = \frac{1}{4}e^{-x} + \frac{3}{4}\left(2e^{-2x}\right)$$

$$\Rightarrow \quad M_X(t) = E\left(e^{tX}\right) = \int_0^{+\infty} e^{tx}\left[\frac{1}{4}e^{-x} + \frac{3}{4}\left(2e^{-2x}\right)\right]dx = \frac{1}{4}\left(\frac{1}{1-t}\right) + \frac{3}{4}\left(\frac{2}{2-t}\right)$$

Problem 11

$$M_X(t) = e^{7t}\left(0.2e^{6t} + 0.8\right)^{10}$$

Calculate $E(X)$ and $Var(X)$

Solution

Method 1

$M_X(t)$ is the product of two terms e^{7t} and $\left(0.2e^{6t} + 0.8\right)^{10}$. e^{7t} is the MGF for the constant 7. $\left(0.2e^{6t} + 0.8\right)^{10}$ is the MGF for $6Y$, where Y is a binomial random variable with parameters $n = 10$ and $p = 0.2$.

So $X = 6Y + 7$.

$$E(X) = E(6Y + 7) = 6E(Y) + 7 = 6(n\,p) + 7 = 6(10 \times 0.2) + 7 = 19$$
$$Var(X) = Var(6Y + 7) = 6^2 Var(Y) = 36(npq) = 36(10 \times 0.2 \times 0.8) = 57.6$$

Method 2

The following standard approach is labor intensive:

$$E(X)=\left[\frac{d}{dt}M_X(t)\right]_{t=0}=\left\{\frac{d}{dt}\left[e^{7t}\left(0.2e^{6t}+0.8\right)^{10}\right]\right\}_{t=0}$$

$$E(X^2)=\left[\frac{d^2}{dt^2}M_X(t)\right]_{t=0}=\left\{\frac{d^2}{dt^2}\left[e^{7t}\left(0.2e^{6t}+0.8\right)^{10}\right]\right\}_{t=0}$$

We'll use the following shortcut:

$$\boxed{E(X)=\left[\frac{d}{dt}\ln M_X(t)\right]_{t=0},\qquad Var(X)=\left[\frac{d^2}{dt^2}\ln M_X(t)\right]_{t=0}}$$

This shortcut is very useful if MGF is the product of several functions of t.

Proof. $$\frac{d}{dt}\ln M_X(t)=\frac{1}{M_X(t)}\left[\frac{d}{dt}M_X(t)\right]$$

$$\left[\frac{d}{dt}\ln M_X(t)\right]_{t=0}=\left[\frac{M_X^{\prime}(t)}{M_X(t)}\right]_{t=0}=\frac{1}{\left[M_X(t)\right]_{t=0}}\left[M_X^{\prime}(t)\right]_{t=0}$$

Because $\left[M_X(t)\right]_{t=0}=1$ is true for any MGF, we have:

$$\left[\frac{d}{dt}\ln M_X(t)\right]_{t=0}=\left[M_X^{\prime}(t)\right]_{t=0}=E(X)$$

$$\frac{d^2}{dt^2}\ln M_X(t)=\frac{d}{dt}\left[\frac{d}{dt}\ln M_X(t)\right]=\frac{d}{dt}\left[\frac{M_X^{\prime}(t)}{M_X(t)}\right]=\frac{M_X^{\prime\prime}(t)M_X(t)-M_X^{\prime}(t)M_X^{\prime}(t)}{\left[M_X(t)\right]^2}$$

$$=\frac{M_X^{\prime\prime}(t)M_X(t)-\left[M_X^{\prime}(t)\right]^2}{\left[M_X(t)\right]^2}$$

$$\Rightarrow\quad\left[\frac{d^2}{dt^2}\ln M_X(t)\right]_{t=0}=\left\{\frac{M_X^{\prime\prime}(t)M_X(t)-\left[M_X^{\prime}(t)\right]^2}{\left[M_X(t)\right]^2}\right\}_{t=0}$$

$$=\frac{M_X^{\prime\prime}(t=0)M_X(t=0)-\left[M_X^{\prime}(t=0)\right]^2}{\left[M_X(t=0)\right]^2}$$

However, $M_X(t=0)=1$, $M_X^{\prime}(t=0)=E(X)$, $M_X^{\prime\prime}(t=0)=E(X^2)$.

$$\Rightarrow \quad \left[\frac{d^2}{dt^2}\ln M_X(t)\right]_{t=0} = E\left(X^2\right) - E^2(X) = Var(X)$$

Back to the problem.

$$M_X(t) = e^{7t}\left(0.2e^{6t}+0.8\right)^{10} \quad \Rightarrow \quad \ln M_X(t) = 7t + 10\ln\left(0.2e^{6t}+0.8\right)$$

$$\frac{d}{dt}\ln M_X(t) = \frac{d}{dt}\left[7t + 10\ln\left(0.2e^{6t}+0.8\right)\right] = 7 + \frac{10(0.2)(6)e^{6t}}{0.2e^{6t}+0.8} = 7 + \frac{12e^{6t}}{0.2e^{6t}+0.8}$$

$$= 7 + \frac{12}{0.2+0.8e^{-6t}}$$

$$\frac{d^2}{dt^2}\ln M_X(t) = \frac{d}{dt}\left(7 + \frac{12}{0.2+0.8e^{-6t}}\right) = 12\frac{d}{dt}\left(0.2+0.8e^{-6t}\right)^{-1}$$

$$= 12(-1)(0.8)(-6)\left(0.2+0.8e^{-6t}\right)^{-2}$$

$$E(X) = \left[\frac{d}{dt}\ln M_X(t)\right]_{t=0} = \left[7 + \frac{12}{0.2+0.8e^{-6t}}\right]_{t=0} = 7 + 12 = 19$$

$$Var(X) = \left[\frac{d^2}{dt^2}\ln M_X(t)\right]_{t=0} = \left[12(-1)(0.8)(-6)\left(0.2+0.8e^{-6t}\right)^{-2}\right]_{t=0} = 57.6$$

Problem 12

A machine has two components. One component is at work. The 2nd component sits idle but is activated after the first one has failed. The machine fails only after both components have failed. Let X, Y represent each component's time until failure. Let Z represents the machine's time until failure.

X, Y have the following moment generating functions:

$$M_X(t) = \frac{1}{1-t}, \qquad M_Y(t) = \frac{1}{1-2t}$$

Find $E(Z), Var(Z), E\left(Z^3\right)$. Assume X, Y are independent.

Solution

$Z = X + Y$

$M_Z(t) = M_X(t) \times M_Y(t) = \dfrac{1}{1-t} \times \dfrac{1}{1-2t}$. To simplify the calculation, set

$$\frac{1}{1-t} \times \frac{1}{1-2t} = \frac{A}{1-2t} + \frac{B}{1-t} = \frac{A(1-t) + B(1-2t)}{(1-2t)(1-t)} = \frac{(A+B) - (A+2B)t}{(1-2t)(1-t)}$$

For $\dfrac{1}{(1-2t)(1-t)} = \dfrac{(A+B)-(A+2B)t}{(1-2t)(1-t)}$ to hold, we must have:

$A + B = 1$, $A + 2B = 0$. Solving these equations, we have: $A = 2$, $B = -1$

$$\Rightarrow \quad \frac{1}{1-t} \times \frac{1}{1-2t} = \frac{2}{1-2t} - \frac{1}{1-t}$$

$$M_Z(t) = M_X(t) \times M_Y(t) = \frac{1}{1-t} \times \frac{1}{1-2t} = \frac{2}{1-2t} - \frac{1}{1-t} = 2(1-2t)^{-1} - (1-t)^{-1}$$

$$\frac{d}{dt} M_Z(t) = \frac{d}{dt}\left[2(1-2t)^{-1} - (1-t)^{-1}\right] = 2^2(1-2t)^{-2} - (1-t)^{-2}$$

$$\frac{d}{dt}\left[\frac{d}{dt} M_Z(t)\right] = \frac{d}{dt}\left[2^2(1-2t)^{-2} - (1-t)^{-2}\right] = 2^4(1-2t)^{-3} - 2(1-t)^{-3}$$

$$\frac{d}{dt}\left[\frac{d^2}{dt^2} M_Z(t)\right] = \frac{d}{dt}\left[2^4(1-2t)^{-3} - 2(1-t)^{-3}\right] = 2^4(3)(2)(1-2t)^{-4} - 2(3)(1-t)^{-4}$$

$$E(Z) = \left[\frac{d}{dt} M_Z(t)\right]_{t=0} = \left[2^2(1-2t)^{-2} - (1-t)^{-2}\right]_{t=0} = 3$$

$$E(Z^2) = \left(\frac{d}{dt}\left[\frac{d}{dt} M_Z(t)\right]\right)_{t=0} = \left[2^4(1-2t)^{-3} - 2(1-t)^{-3}\right]_{t=0} = 14$$

$$Var(Z) = E(Z^2) - E^2(Z) = 14 - 3^2 = 5$$

$$E(Z^3) = \left[\frac{d^3}{dt^3} M_Z(t)\right]_{t=0} = \left[2^4(3)(2)(1-2t)^{-4} - 2(3)(1-t)^{-4}\right]_{t=0} = 90$$

If you have memorized the moment generating function of exponential distribution, you will notice that X is exponentially distributed with a mean of 1 and Y is exponentially distributed with a mean of 2. Then

$$E(Z)=E(X+Y)=E(X)+E(Y)=1+2=3$$

$$Var(Z)=Var(X+Y)=Var(X)+Var(Y)=1^2+2^2=5$$

Though we can quickly calculate $E(Z)$ and $Var(Z)$ without using the moment generating function, we cannot easily calculate $E\left(Z^3\right)$ this way. Now you see the power of the moment generating function.

Problem 13

X, Y are two independent random variables with the following moment generating functions:

$$M_X(t)=e^{-t+t^2}, \quad M_Y(t)=e^{2t+\frac{t^2}{2}}$$

Let $Z=X-5Y$. Find $Var(Z)$.

Solution

If you have memorized the moment generating function of a normal variable X :

$$M_X(t)=e^{\mu t+\frac{\sigma^2 t^2}{2}}$$

you can then solve this problem quickly.

In this problem, X is normal with a mean of -1 and a variance of 2; Y is normal with a mean of 2 and a variance of 1.

Then $Var(Z)=Var(X-5Y)=Var(X)+5^2Var(Y)=2+5^2(1)=27$

If you have not memorized the moment generating function for normal distribution or for any other distribution:

$$M_{X-5Y}\left(t\right)=M_X\left(t\right)\times M_{-5Y}\left(t\right)=M_X\left(t\right)\times M_Y\left(-5t\right)$$

$$M_X\left(t\right)=e^{-t+t^2}, \quad M_{-5Y}\left(t\right)=M_Y\left(-5t\right)=e^{2(-5t)+\frac{(-5t)^2}{2}}$$

$$M_{X-5Y}(t)=e^{-t+t^2}\times e^{2(-5t)+\frac{(-5t)^2}{2}}=e^{-t+t^2+2(-5t)+\frac{(-5t)^2}{2}}=e^{-11t+13.5t^2}$$

$$\frac{d}{dt}M_{X-5Y}\left(t\right)=\frac{d}{dt}e^{-11t+13.5t^2}=\left(-11+27t\right)e^{-11t+13.5t^2}$$

$$\frac{d^2}{dt^2}M_{X-5Y}\left(t\right)=\frac{d}{dt}\left[\left(-11+27t\right)e^{-11t+13.5t^2}\right]=\left[\left(-11+27t\right)^2+27\right]e^{-11t+13.5t^2}$$

$$E\left(X-5Y\right)=\frac{d}{dt}M_{X-5Y}\left(t\right)\bigg|_{t=0}=\left(-11+27t\right)e^{-11t+13.5t^2}\bigg|_{t=0}=-11$$

$$E\left[(X-5Y)^2\right]=\frac{d^2}{dt^2}M_{X-5Y}(t)\bigg|_{t=0}$$

$$=\left\{\left[\left(-11+27t\right)^2+27\right]e^{-11t+13.5t^2}\right\}_{t=0}=\left(-11\right)^2+27$$

$$Var(X-5Y)=E\left[(X-5Y)^2\right]-\left[E(X-5Y)\right]^2=27$$

Alternatively:

$$M_{X-5Y}\left(t\right)=e^{-11t+13.5t^2}$$

$$\ln M_{X-5Y}\left(t\right)=-11t+13.5t^2$$

$$\frac{d}{dt}\ln M_{X-5Y}\left(t\right)=\frac{d}{dt}\left(-11t+13.5t^2\right)=-11+27t$$

$$\frac{d^2}{dt^t}\ln M_{X-5Y}\left(t\right)=\frac{d}{dt}\left(-11+27t\right)=27$$

$$\Rightarrow\quad E\left(X-5Y\right)=\left[\frac{d}{dt}\ln M_{X-5Y}\left(t\right)\right]_{t=0}=-11$$

$$\Rightarrow\quad Var\left(X-5Y\right)\left[=\frac{d^2}{dt^2}\ln M_{X-5Y}\left(t\right)\right]_{t=0}=\left[27\right]_{t=0}=27$$

You see that the shortcut really cuts to the chase.

Q14

Random variable X has the following moment generating function:

$$M_X(t)=\frac{1}{2}e^t\left(1+e^{2t}\right)$$

Calculate $Var(X)$.

Solution

Method 1

$$M_X(t)=\frac{1}{2}e^t\left(1+e^{2t}\right)=\frac{1}{2}e^t+\frac{1}{2}e^{3t}$$

e^t is the MGF for a constant of 1; e^{3t} is the MGF for a constant of 3. Remember $M_b(t)=e^{bt}$.

So we see that X takes on the value of 1 and 3 each with probability of $\frac{1}{2}$:

$$X=\begin{cases}1 & \text{with probability of 0.5}\\ 3 & \text{with probability of 0.5}\end{cases}$$

$$\Rightarrow \quad E(X)=0.5(1+3)=2,\ E\left(X^2\right)=0.5\left(1^2+3^2\right)=5,\ Var(X)=5-2^2=1$$

Method 2

$$M_X(t)=\frac{1}{2}e^t\left(1+e^{2t}\right)=\frac{1}{2}\left(e^t+e^{3t}\right)$$

$$\frac{d}{dt}M_X(t)=\frac{1}{2}\frac{d}{dt}\left[e^t+e^{3t}\right]=\frac{1}{2}\left(e^t+3e^{3t}\right)$$

$$\frac{d^2}{dt^2}M_X(t)=\frac{d}{dt}\left[\frac{1}{2}\left(e^t+3e^{3t}\right)\right]=\frac{1}{2}\left(e^t+9e^{3t}\right)$$

$$E(X)=\left[\frac{d}{dt}M_X(t)\right]_{t=0}=\left[\frac{1}{2}\left(e^t+3e^{3t}\right)\right]_{t=0}=\frac{1}{2}(1+3)=2$$

$$E\left(X^2\right)=\left[\frac{d^2}{dt^2}M_X(t)\right]_{t=0}=\left[\frac{1}{2}\left(e^t+9e^{3t}\right)\right]_{t=0}=\frac{1}{2}(1+9)=5$$

$$Var(X)=E\left(X^2\right)-E^2(X)=5-2^2=1$$

Method 3

$$\ln M_X(t)=\ln\left[\frac{1}{2}e^t\left(1+e^{2t}\right)\right]=\ln\frac{1}{2}+t+\ln\left(1+e^{2t}\right)$$

$$\frac{d}{dt}\ln M_X(t)=\frac{d}{dt}\left[\ln\frac{1}{2}+t+\ln\left(1+e^{2t}\right)\right]=1+\frac{2e^{2t}}{1+e^{2t}}=1+\frac{2}{e^{-2t}+1}=1+2\left(e^{-2t}+1\right)^{-1}$$

$$\frac{d^2}{dt^2}\ln M_X(t)=\frac{d}{dt}\left[1+2\left(e^{-2t}+1\right)^{-1}\right]=2(-1)\left(e^{-2t}+1\right)^{-2}(-2)e^{-2t}$$

$$Var(X)=\left[\frac{d^2}{dt^2}\ln M_X(t)\right]_{t=0}=\left[2(-1)\left(e^{-2t}+1\right)^{-2}(-2)e^{-2t}\right]_{t=0}=4(2)^{-2}=1$$

Q15 True/False

Two random variables X and Y have identical moment generating function: $M_X(t)=M_Y(t)=\dfrac{1}{1-t}$. This means that $X=Y$.

Solution False

$\dfrac{1}{1-t}$ is the moment generating function of an exponential random variable with mean 1.

$M_X(t)=M_Y(t)=\dfrac{1}{1-t}$ means that X and Y have the following the same pdf:

$$f_X(x)=e^{-x},\quad f_X(y)=e^{-y}$$

However, having the same probability distribution function doesn't mean that two random variables are the same.

To bring this concept home, let's consider two unbiased coins A and B.

Let

X represent the number of heads we get if we flip coin A once
Y represent the number of heads we get if we flip coin B once

Clearly, X and Y have identical probability mass function:

$$X=\begin{cases}1 & \text{with probability } 0.5\\ 0 & \text{with probability } 0.5\end{cases} \qquad Y=\begin{cases}1 & \text{with probability } 0.5\\ 0 & \text{with probability } 0.5\end{cases}$$

X and Y have identical MGF:

$$M_X(t)=M_Y(t)=E\left(e^{tX}\right)=E\left(e^{tY}\right)=0.5\left(e^{t\times 0}+e^{t\times 1}\right)=0.5\left(1+e^t\right)$$

However, it's not true that $X=Y$. For example, if you flip Coin A, you may get a head (so $X=1$). If you flip Coin B, you may get a tail (so $Y=0$).

Key point to remember:

If two random variables X and Y have the same probability function or the same MGF, it doesn't mean $X=Y$.

Problem 16

An exam candidate was solving the following problem (Problem 12 in this chapter):

A machine has two components. One component is at work. The 2nd component sits idle but is activated after the first one has failed. The machine fails only after both components have failed. Let X, Y represent each component's time until failure. Let Z represents the machine's time until failure.

X, Y have the following moment generating functions:

$$M_X(t)=\frac{1}{1-t}, \qquad M_Y(t)=\frac{1}{1-2t}$$

Find $E(Z), Var(Z), E\left(Z^3\right)$. Assume X, Y are independent.

This is his approach:

X is an exponential random variable with mean 1.

Using the formula $M_{aX}(t)=M_X(at)$:

$$M_{2X}(t)=M_X(2t)=\frac{1}{1-2t}. \quad \Rightarrow \quad M_Y(t)=M_{2X}(t)=\frac{1}{1-2t} \quad \Rightarrow Y=2X$$

$$Z=X+Y=X+2X=3X$$

$$\Rightarrow \; E(Z)=E(3X)=3E(X)=3$$

$\Rightarrow\ Var(Z) = Var(3X) = 9Var(X) = 9$

$\Rightarrow\ E(Z^3) = E(27X^3) = 27E(X^3) = 27\int_0^\infty x^3 e^{-x} dx$

After a bunch of integration by parts: $\int_0^\infty x^3 e^{-x} dx = 6$

$\Rightarrow\ E(Z^3) = E(27X^3) = 27E(X^3) = 27\int_0^\infty x^3 e^{-x} dx = 27(6) = 162$

Explain why this approach is wrong.

Solution

This mistake is in this step:

$$M_Y(t) = M_{2X}(t) = \frac{1}{1-2t} \quad \Rightarrow Y = 2X$$

As explained before, two random variables having the same MGF doesn't mean these two random variables are the same. It's not true that $Y = 2X$.

Homework for you: #35, May 2000; #11, #27, Nov 2000

Chapter 34 Joint moment generating function

The joint moment generating function of (X, Y) is defined as

$$M_{X,Y}(s,t) = E\left[e^{sx+ty}\right]$$

If X, Y are discrete, then

$$M_{X,Y}(s,t) = E\left(e^{sx+ty}\right) = \sum_{x=-\infty}^{+\infty} \sum_{y=-\infty}^{+\infty} e^{sx+ty} f_{X,Y}(x,y)$$

If X, Y are continuous, then

$$M_{X,Y}(s,t) = E\left[e^{sx+ty}\right] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{sx+ty} f_{X,Y}(x,y)\,dxdy$$

Let's focus on the case when X, Y are continuous. If the joint MGF is ever tested, the continuous joint MGF is more likely to be tested.

Important properties of joint MGF:

(1) If $M_{X,Y}(s,t)$ is finite in a rectangle containing $(0,0)$, then the joint pdf $f_{X,Y}(x,y)$ is completely determined by $M_{X,Y}(s,t)$.

(2) $M_{X,Y}(s, 0) = M_X(s)$. By setting $t = 0$, the joint MGF becomes the X marginal MGF.

(3) $M_{X,Y}(0, t) = M_Y(t)$. By setting $s = 0$, the joint MGF becomes the Y marginal MGF.

(4) $M_{X,Y}(s,t) = M_{X+Y}(t)$. If we set $s = t$, then the joint MGF becomes $M_{X+Y}(t)$, the MGF of $X + Y$.

(5) $E\left(X^n Y^m\right) = \dfrac{\partial^{n+m}}{\partial^n s \partial^m t} M_{X,Y}(s,t)\Big|_{s=t=0}$

Once you know the definition of a joint MGF, the remaining work is doing double integration. You'll need to apply the process described in Chapter 26 to quickly and correctly complete the double integration.

Course 1 May 2003 #39 is the only problem about the joint MGF.

Problem 1

Prove that $E\left(X^n Y^m\right)=\dfrac{\partial^{n+m}}{\partial^n s \partial^m t} M_{X,Y}(s,t)\Big|_{s=t=0}$.

Solution

$$\boxed{\text{If } n=1 \text{ and } m=0: \quad E(X)=\frac{d}{ds} M_{X,Y}(s,t)\Big|_{s=t=0}.}$$

$$\frac{d}{ds} M_{X,Y}(s,t)=\frac{d}{ds} E\left[e^{sx+ty}\right]=E\left[\frac{d}{ds}\left(e^{sx+ty}\right)\right]=E\left[xe^{sx+ty}\right]$$

(In the above formula, x, y are treated as constants)

$$\Rightarrow \left[\frac{d}{ds} M_{X,Y}(s,t)\right]_{s=t=0}=E\left[\frac{d}{ds}\left(e^{sx+ty}\right)\right]_{s=t=0}=E\left[xe^0\right]=E(X)$$

$$\boxed{\text{If } n=0 \text{ and } m=1: \quad E(Y)=\frac{d}{dt} M_{X,Y}(s,t)\Big|_{s=t=0}.}$$

$$\frac{d}{dt} M_{X,Y}(s,t)=\frac{d}{dt} E\left[e^{sx+ty}\right]=E\left[\frac{d}{dt}\left(e^{sx+ty}\right)\right]=E\left[ye^{sx+ty}\right]$$

(In the above formula, x, y are treated as constants)

$$\Rightarrow \left[\frac{d}{dt} M_{X,Y}(s,t)\right]_{s=t=0}=E\left[\frac{d}{dt}\left(e^{sx+ty}\right)\right]_{s=t=0}=E\left[ye^0\right]=E(Y)$$

$$\boxed{\text{If } n=m=1: \quad E(XY)=\frac{\partial^2}{\partial s \partial t} M_{X,Y}(s,t)\Big|_{s=t=0}.}$$

$$\frac{\partial^2}{\partial s \partial t} M_{X,Y}(s,t)=\frac{\partial^2}{\partial s \partial t} E\left[e^{sx+ty}\right]=\frac{\partial}{\partial s}\left\{\frac{\partial}{\partial t} E\left[e^{sx+ty}\right]\right\}=\frac{\partial}{\partial s} E\left[\frac{\partial}{\partial t}\left(e^{sx+ty}\right)\right]$$

$$=\frac{\partial}{\partial s} E\left[\frac{\partial}{\partial t}\left(e^{sx+ty}\right)\right]=\frac{\partial}{\partial s} E\left[\left(ye^{sx+ty}\right)\right]=E\left[\frac{\partial}{\partial s}\left(ye^{sx+ty}\right)\right]=E\left[\left(xye^{sx+ty}\right)\right]$$

(In the above formula, x, y are treated as constants)

$$\Rightarrow \frac{\partial^2}{\partial s \partial t} M_{X,Y}(s,t)\Big|_{s=t=0}=\left[E\left(xye^{sx+ty}\right)\right]_{s=t=0}=E(XY)$$

If $n=2$ and $m=0$: $E\left(X^2\right)=\dfrac{\partial^2}{\partial s^2}M_{X,Y}\left(s,t\right)\Big|_{s=t=0}$.

$$\frac{\partial^2}{\partial s^2}M_{X,Y}\left(s,t\right)=\frac{\partial}{\partial s}\left[\frac{\partial}{\partial s}M_{X,Y}\left(s,t\right)\right]=\frac{\partial}{\partial s}\left[\frac{\partial}{\partial s}E\left(e^{sx+ty}\right)\right]=\frac{\partial}{\partial s}\left[E\left(xe^{sx+ty}\right)\right]=E\left(x^2e^{sx+ty}\right)$$

(In the above formula, x, y are treated as constants)

$$\Rightarrow \quad \frac{\partial^2}{\partial s^2}M_{X,Y}\left(s,t\right)\Big|_{s=t=0}=\left[E\left(x^2e^{sx+ty}\right)\right]_{s=t=0}=E\left(X^2\right)$$

If $n=0$ and $m=2$: $E\left(Y^2\right)=\dfrac{\partial^2}{\partial t^2}M_{X,Y}\left(s,t\right)\Big|_{s=t=0}$.

$$\frac{\partial^2}{\partial t^2}M_{X,Y}\left(s,t\right)=\frac{\partial}{\partial t}\left[\frac{\partial}{\partial t}M_{X,Y}\left(s,t\right)\right]=\frac{\partial}{\partial t}\left[\frac{\partial}{\partial t}E\left(e^{sx+ty}\right)\right]=\frac{\partial}{\partial t}\left[E\left(ye^{sx+ty}\right)\right]=E\left(y^2e^{sx+ty}\right)$$

(In the above formula, x, y are treated as constants)

$$\Rightarrow \quad \frac{\partial^2}{\partial t^2}M_{X,Y}\left(s,t\right)\Big|_{s=t=0}=\left[E\left(y^2e^{sx+ty}\right)\right]_{s=t=0}=E\left(Y^2\right)$$

Following this line of reasoning, we see that

$$E\left(X^nY^m\right)=\frac{\partial^{n+m}}{\partial^n s\partial^m t}M_{X,Y}\left(s,t\right)\Big|_{s=t=0}$$

Problem 2

$\left(X,Y\right)$ is uniformly distributed over $0<y<x<1$. Find $M_{X,Y}\left(s,t\right)$.

Solution

All you need to do is to find

$$E\left[e^{sX+tY}\right]=\iint\limits_{0<y<x<1}e^{sX+tY}f_{X,Y}\left(x,y\right)dxdy$$

To complete this double integration, you'll use the procedure described in Chapter 26:

- Determine 2-D region.
- Set up the outer integration
- Set up the inner integration
- Evaluate the double integration

I won't draw the 2-D graph for you or show you the other steps; I'll just give you the result. However, you might want work step by step. Make sure you understand how to do this double integration.

To complete this double integration, we need to find the joint pdf. Because (X,Y) is uniformly distributed over $0<y<x<1$, we have

$f_{X,Y}(x,y)=k$ where k is a positive constant.

$$\iint_{0<y<x<1} f_{X,Y}(x,y)dxdy=1 \quad \Rightarrow \quad k \iint_{0<y<x<1} dxdy=1$$

$$\iint_{0<y<x<1} dxdy = Area \text{ of } (0<y<x<1)=\frac{1}{2} \quad \Rightarrow \quad k=2$$

$$E\left[e^{sx+ty}\right]= \iint_{0<y<x<1} e^{sx+ty} f_{X,Y}(x,y)dxdy = 2 \iint_{0<y<x<1} e^{sx+ty}dxdy$$

$$\iint_{0<y<x<1} e^{sx+ty}dxdy = \int_0^1\int_0^x e^{sx+ty}dydx = \int_0^1 e^{sx}\left[\int_0^x e^{ty}dy\right]dx$$

$$= \int_0^1 e^{sx}\frac{1}{t}\left(e^{xt}-1\right)dx = \frac{1}{t}\left[\frac{1}{s+t}\left(e^{t+s}-1\right)-\frac{1}{s}\left(e^{s}-1\right)\right]$$

$$\Rightarrow \quad M_{X,Y}(s,t)=\frac{2}{t}\left[\frac{1}{s+t}\left(e^{t+s}-1\right)-\frac{1}{s}\left(e^{s}-1\right)\right] \quad \text{where } s\neq 0, t\neq 0$$

Problem 3

(X,Y) has the following pdf:

$f_{X,Y}(x,y)=ke^{-x-2y}$ where $0<x<y<\infty$

Find $M_{X,Y}(s,t)$.

Solution

First, we need to find k.

$$\iint_{0<x<y<\infty} ke^{-x-2y}dxdy = k \iint_{0<x<y<\infty} e^{-x-2y}dxdy = 1$$

$$\iint_{0<x<y<\infty} e^{-x-2y}\,dxdy = \int_0^\infty \int_0^y e^{-x-2y}\,dxdy = \int_0^\infty e^{-2y}\left[\int_0^y e^{-x}\,dx\right]dy = \int_0^\infty e^{-2y}\left[e^{-y}-1\right]dy = \frac{1}{6}$$

$$\Rightarrow \quad k = 6$$

$$E\left[e^{sx+ty}\right] = \iint_{0<x<y<\infty} 6e^{sx+ty}e^{-x-2y}\,dxdy = \iint_{0<x<y<\infty} 6e^{(s-1)x}e^{(t-2)y}\,dxdy = \int_0^\infty \int_0^y 6e^{(s-1)x}e^{(t-2)y}\,dxdy$$

$$= \int_0^\infty e^{(t-2)y}\int_0^y 6e^{(s-1)x}\,dxdy = 6\int_0^\infty e^{(t-2)y}\,\frac{1}{s-1}\left[e^{(s-1)y}-1\right]dy \quad \text{(where } s \neq 1\text{)}$$

$$= \frac{6}{s-1}\left\{\int_0^\infty e^{(s+t-3)y}\,dy - \int_0^\infty e^{(t-2)y}\,dy\right\}$$

$$= \frac{6}{s-1}\left\{\int_0^\infty e^{-(3-s-t)y}\,dy - \int_0^\infty e^{-(2-t)y}\,dy\right\}$$

$$= \frac{6}{s-1}\left[\frac{1}{3-s-t} - \frac{1}{2-t}\right] \quad \text{(where } s+t<3,\ t<2,\ s\neq 1\text{)}$$

$$= \frac{6}{s-1}\left[\frac{1}{t-2} - \frac{1}{s+t-3}\right]$$

$$\Rightarrow \quad M_{X,Y}(s,t) = E\left[e^{s\,x+t\,y}\right] = \frac{6}{s-1}\left[\frac{1}{t-2} - \frac{1}{s+t-3}\right] \text{ where } s+t<3,\ t<2,\ s\neq 1$$

Problem 4

Random variables X and Y have the following joint pdf:

$f(x,y) = k$ where $0 \le x \le 1$, $0 \le y \le 1$, and $x+y \le 1$.

Find the joint moment generating function of X and Y.
Find the moment generating function for X using the joint MGF.
Find the moment generating function for Y using the joint MGF.

Solution

First, we need to find the constant k.

$$\iint_A f(x,y)\,dxdy = 1$$

Where A represents the area $0 \le x \le 1$, $0 \le y \le 1$, and $x + y \le 1$.

$$\iint_A f(x,y)\,dxdy = 1, \quad \Rightarrow \quad k\iint_A dxdy = 1,$$

Since $\iint_A dxdy = \text{Area of A}$, we have $k = \dfrac{1}{\text{Area of A}}$

You need to memorize the above result. That is, if X and Y are uniformly distributed over an area, then the joint pdf is: $f(x,y) = \dfrac{1}{\text{Area}}$

If you draw the 2-D diagram for $0 \le x \le 1$, $0 \le y \le 1$, and $x + y \le 1$, you'll find that the area is 0.5. Then the joint pdf is simply $\dfrac{1}{0.5} = 2$.

Next, we'll find the joint MGF.

$$M_{X,Y}(s,t) = 2\int_0^1\int_0^{1-x} e^{sx+ty}\,dy\,dx = 2\int_0^1 e^{sx}\left(\int_0^{1-x} e^{ty}dy\right)dx = 2\int_0^1 e^{sx}\frac{1}{t_2}\left[e^{t(1-x)} - 1\right]dx$$

$$= \frac{2}{t}\int_0^1\left[e^{(s-t)x}e^t - e^{sx}\right]dx = \frac{2}{s\,t(s-t)}\left[t\,e^s - s\,e^t + (s-t)\right]$$

If we want to find the MGF for X or Y, we don't need to start from scratch. We can use $M_{X,Y}(s,t)$.

$$M_X(s) = M_{X,Y}(s,t=0) = \lim_{t\to 0}\frac{2}{s\,t(s-t)}\left[t\,e^s - s\,e^t + (s-t)\right]$$

Notice that $\left[t\,e^s - s\,e^t + (s-t)\right] \to 0$ and $s\,t(s-t) \to 0$. So we have $\dfrac{0}{0}$; we'll need to use I'Hospital's rule.

$$\frac{d}{dt}2\left[t\,e^s - s\,e^t + (s-t)\right] = 2\left(e^s - se^t - 1\right)$$

$$\frac{d}{dt}\left[s\,t\left(s-t\right)\right]=s\left(s-2t\right)$$

$$\lim_{t\to 0}\frac{2}{s\,t\left(s-t\right)}\left[t\,e^{s}-s\,e^{t}+\left(s-t\right)\right]=\lim_{t\to 0}\frac{2\left(e^{s}-s\,e^{t}-1\right)}{s\left(s-2t\right)}=\frac{2\left(e^{s}-s-1\right)}{s^{2}}$$

$$M_{X}\left(s\right)=M_{X,Y}\left(s,t=0\right)=\frac{2\left(e^{s}-s-1\right)}{s^{2}}$$

Similarly,

$$M_{Y}\left(t\right)=M_{X,Y}\left(s=0,t\right)=\frac{2\left(e^{t}-t-1\right)}{t^{2}}$$

Chapter 35 Markov's inequality, Chebyshev inequality

This topic is not on the SOA syllabus. However, SOA annoyingly tested Chebyshev Inequality in May and September 2005. So you might want to learn this topic.

Please note that Markov's inequality and Chebyshev Inequality are almost never used to estimate probability for any real world applications. Given today's computing power, it's never necessary to estimate probability using Markov's inequality and Chebyshev Inequality.

Markov's inequality and Chebyshev Inequality are important only theoretically. They give us a crude estimate of probability.

Now let's move on to the topic.

Markov's inequality

If a non-negative random variable X (i.e. $X \geq 0$) has a finite mean (i.e. if $E(X)$ exists), then

$$P(X \geq a) \leq \frac{E(X)}{a}, \text{ for all } a > 0$$

Markov's inequality really says that if a non-negative random variable has a finite mean, then chances are slim that it will take on a huge value.

It's easy to prove Markov's inequality. We'll prove Markov's inequality when X is a continuous non-negative random variable. The proof is similar if X is a discrete non-negative random variable.

Proof:

$$E(X) = \int_0^{+\infty} x\, f(x)dx = \int_0^a x\, f(x)dx + \int_a^{+\infty} x\, f(x)dx \quad \geq \quad \int_a^{+\infty} x\, f(x)dx$$

$$\int_a^{+\infty} x\, f(x)dx \geq \int_a^{+\infty} af(x)dx = a\int_a^{+\infty} f(x)dx = a\, P(X \geq a)$$

$$\Rightarrow \quad E(X) \quad \geq \quad a\, P(X \geq a), \quad \Rightarrow \quad P(X \geq a) \leq \frac{E(X)}{a}$$

Markov's inequality is really simple. The trouble, however, is to memorize the direction of the inequality. It's easy to get confused in the exam and use a wrong formula such as

$E(X) \quad \geq \quad a\, P(X \leq a)$ or $P(X \geq a) \geq \dfrac{E(X)}{a}$.

You have two ways to memorize Markov's inequality: one way is to memorize $E(X) \geq a\,P(X \geq a)$; the other is to memorize $P(X \geq a) \leq \dfrac{E(X)}{a}$.

Method 1: Memorize $E(X) \geq a\,P(X \geq a)$.

To memorize $E(X) \geq a\,P(X \geq a)$, remember two rules:

Rule 1: The two inequality symbols both point to the right. In other words, you need to use "$\geq$" and "$\geq$."

Don't write:

$E(X) \geq a\,P(X \leq a)$ (Wrong. Here the 1st equality symbol points to the right; the 2nd points to the left.)

$E(X) \leq a\,P(X \geq a)$ (Wrong. Here the 1st equality symbol points to the left; the 2nd points to the right.)

$E(X) \leq a\,P(X \leq a)$ (Wrong. Here the two equality symbols both point to the left.)

Rule 2: If we cut off $X \in [0, a]$ and use $X \geq a$ to calculate the mean $E(X)$, we'll get a lower bound of the mean." So we should have the expression "$E(X) \geq$ something," not "$E(X) \leq$ something."

This is why we have

$$E(X) \geq a\,P(X \geq a)$$

And we don't have

$$E(X) \leq a\,P(X \geq a) \quad \text{(Wrong !)}$$

Method 2: Memorize $P(X \geq a) \leq \dfrac{E(X)}{a}$.

How to memorize:

The 1st inequality symbol points to the right; the 2nd inequality symbol points to the left; the letter "a" is trapped in between. The letter "a" is being pointed from both sides. In other words, we need to write "$\geq a \leq$."

This way, you won't write:

$P(X \geq a) \geq \dfrac{E(X)}{a}$ (Wrong! Two inequality symbols point to the right.)
$P(X \leq a) \leq \dfrac{E(X)}{a}$ (Wrong! Two inequality symbols point to the left.)
$P(X \leq a) \geq \dfrac{E(X)}{a}$ (Wrong! Two inequality symbols point to the same direction.)

To get a feel of the formula $P(X \geq a) \leq \dfrac{E(X)}{a}$, we'll let $a \to 0$. If $a \to 0$, then $\dfrac{E(X)}{a} \to +\infty$ because $E(X)$ is finite. Since $P(X \geq a) \leq 1$, evidently we can't have $P(X \geq a) \geq \dfrac{E(X)}{a} \to +\infty$. So we'll need to write $P(X \geq a) \leq \dfrac{E(X)}{a}$.

Hope by now you can write out Markov's inequality correctly. Next, let's move onto Chebyshev Inequality.

Chebyshev Inequality

If a random variable X has mean μ and variance σ^2, then $P(\lvert X-\mu\rvert \geq c) \leq \dfrac{\sigma^2}{c^2}$, for all $c > 0$

We can easily prove Chebyshev Inequality using Markov Inequality.
Consider a new random variable $Y = (X-\mu)^2$. Notice that Y is always non-negative.
Using Markov Inequality and setting $a = c^2$, we have:

$$E(Y) \geq c^2 P(Y \geq c^2)$$

$$E(Y) = E(X-\mu)^2 = Var(X)$$

$$P(Y \geq c^2) = P\left[(X-\mu)^2 \geq c^2\right] = P(\lvert X-\mu\rvert \geq c)$$

Then it follows:

$$Var(X) \geq c^2 P(\lvert X-\mu\rvert \geq c), \quad \Rightarrow \quad P(\lvert X-\mu\rvert \geq c) \leq \frac{Var(X)}{c^2}$$

To memorize $P(|X-\mu| \geq c) \quad \leq \quad \frac{Var(X)}{c^2}$, remember that the letter "c" is being pointed from both sides. In other words, we need to write "$\geq c \leq$."

There's another common expression for Chebyshev Inequality. If we let $c = k\sigma$, where k is a positive constant and σ is the standard deviation of X, then

$$P(|X-\mu| \geq k\sigma) \quad \leq \quad \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

2nd expression of Chebyshev Inequality

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

The probability is at most $\frac{1}{k^2}$ that X is k standard deviation away from its mean.

Problem 1

X is uniformly distributed over $[-1,1]$. Calculate $P(|X| \geq a)$ where $a > 0$ using the following 3 methods:

- The exact method
- Markov Inequality
- Chebyshev Inequality

Solution

The exact method

X is uniformly distributed over $[-1,1]$. The pdf is $f(x) = \frac{1}{1-(-1)} = \frac{1}{2}$. The cdf is $P(X \leq x) = \frac{x+1}{2}$ where $-1 \leq x \leq 1$.

$$P(|X| \geq a) = 1 - P(|X| < a) = 1 - P(|X| \leq a)$$

Please note that $|X|$ is continuous. As a result, $P\left(|X| \geq a\right) = P\left(|X| > a\right)$ and $P\left(|X| < a\right) = P\left(|X| \leq a\right)$.

$$P\left(|X| \leq a\right) = P\left(-a \leq X \leq a\right) = P\left(-a < X \leq a\right) = F\left(a\right) - F\left(-a\right) = \frac{a+1}{2} - \frac{-a+1}{2} = a$$

$$\Rightarrow \quad P\left(|X| \geq a\right) = 1 - P\left(|X| \leq a\right) = 1 - a \text{ where } a \leq 1$$

If $a > 1$, then $P\left(|X| \geq a\right) = 0$

<u>Markov Inequality</u>

If we cut off $|X| \in [0, a]$ when calculating the mean $E\left(|X|\right)$, we'll get a lower bound of $E\left(|X|\right)$.

$$\Rightarrow \quad E\left(|X|\right) \geq aP\left(|X| \geq a\right), \quad \Rightarrow \quad P\left(|X| \geq a\right) \leq \frac{E\left(|X|\right)}{a}$$

Applying the general formula $E\left[y(x)\right] = \int_{-\infty}^{+\infty} y(x) f(x) dx$ and setting $y(x) = |X|$, we have

$$E\left(|X|\right) = \int_{-1}^{1} |X| \, f(x) dx = \frac{1}{2}\int_{-1}^{1} |X| \, dx = \frac{1}{2}\left[\int_{-1}^{0} -x \, dx + \int_{0}^{1} x \, dx\right] = \frac{1}{2}$$

$$\Rightarrow \quad P\left(|X| \geq a\right) \leq \frac{E\left(|X|\right)}{a} = \frac{1}{2a}$$

If $a \leq \frac{1}{2}$, then $\frac{1}{2a} \leq 1$ and $P\left(|X| \geq a\right) \leq \frac{1}{2a} \leq 1$. It doesn't give us any new knowledge. Even without knowing Markov Inequality, we know $P\left(|X| \geq a\right) \leq 1$; the probability of anything must not exceed one.

<u>Chebyshev Inequality</u>

$$P\left(\left|X - E(X)\right| \geq a\right) \leq \frac{Var(X)}{a^2}$$

X is uniformly distributed over $[-1,1]$. We have:

$$E(X)=0,\ Var(X)=\left(\frac{2}{2\sqrt{3}}\right)^2=\frac{1}{3}$$

The general formula is:

If the random variable X is uniformly distributed over $[a, b]$ where $b > a$, then

$$f(x)=\frac{1}{b-a},\ E(X)=\frac{b+a}{2},\ \sigma_X=\frac{b-a}{2\sqrt{3}}$$

$$P\left(\left|X-E(X)\right|\geq a\right)\leq\frac{Var(X)}{a^2}$$

$$\Rightarrow\quad P\left(\left|X\right|\geq a\right)\leq\frac{1}{3a^2}$$

If $a\geq\frac{1}{\sqrt{3}}$, then $3a^2\geq1$ and $\frac{1}{3a^2}\leq1$. Then $P\left(\left|X\right|\geq a\right)\leq\frac{1}{3a^2}\leq1$. This doesn't tell us anything new.

Problem 2

Using Chebyshev Inequality, find k that will guarantee that the probability is 0.95 that the deviation of X from $E(X)$ is no more than $k\sigma$.

Solution

"find k that will guarantee that the probability is 0.95 that the deviation of X from $E(X)$ is no more than $k\sigma$" means the following:

$$P\left[\left|X-E(X)\right|\leq k\sigma\right]\leq0.95$$

$$\Rightarrow\quad P\left[\left|X-E(X)\right|>k\sigma\right]=1-P\left[\left|X-E(X)\right|\leq k\sigma\right]\geq0.95$$

$$\Rightarrow\quad P\left[\left|X-E(X)\right|>k\sigma\right]\leq0.05$$

Since $P\left[\left|X-E(X)\right|>k\sigma\right]=P\left[\left|X-E(X)\right|\geq k\sigma\right]-P\left[\left|X-E(X)\right|=k\sigma\right]$, we'll always have $P\left[\left|X-E(X)\right|>k\sigma\right]\leq P\left[\left|X-E(X)\right|\geq k\sigma\right]$.

So if we can satisfy $P\left[\left|X-E(X)\right|\geq k\sigma\right]\leq 0.05$, we'll guarantee that $P\left[\left|X-E(X)\right|>k\sigma\right]\leq 0.05$. Mathematically, this is:

$$P\left[\left|X-E(X)\right|>k\sigma\right]\leq P\left[\left|X-E(X)\right|\geq k\sigma\right]\leq 0.05$$

Based on Chebyshev Inequality, we have:

$$P\left[\left|X-E(X)\right|\geq k\sigma\right]\leq \frac{Var(X)}{(k\sigma)^2}=\frac{\sigma^2}{(k\sigma)^2}=\frac{1}{k^2}$$

So we need to make sure $\frac{1}{k^2}\leq 0.05$. This gives us $k\geq\sqrt{20}$.

Problem 3

After SOA switched from the pencil-and-paper testing to the computer based testing (CBT) for Exam P, some candidates support CBT and other disapprove. Those who support CBT claim that CBT can be offered more than the pencil-and-paper testing. Those who disapprove CBT cite that they have to pay higher exam fees to take CBT; they have to drive longer to find a CBT center; that computer screens often freeze during the exam.

SOA asks you to find the supporting ratio (i.e. the % of the Exam P candidates who favor CBT) of the Exam P candidate's population. You know that the Exam P population is huge (over 6,000 people can take one exam) and that there's no way you can find the true supporting ratio of the entire Exam P candidate population. As a result, you decide to randomly survey some Exam P candidates and find the supporting ratio of your sample. You are thinking of using the supporting ratio of our sample as an estimate of the true supporting ratio of the population.

SOA wants to ensure that chances are more than 95% that your sample mean will not differ from the true mean of the population by more than 0.01. How many Exam P candidates do you need to survey?

Solution

Let p represent the true supporting ration of the Exam P candidate population. Let $\overline{p}$ represent the supporting ratio of the sample.

Suppose we sample n Exam P candidates. Of these n candidates surveyed, m candidates support CBT. Then $\overline{p}=\frac{m}{n}$. Then m is a binomial distribution with parameter n and p. $E(m)=np$ and $Var(m)=npq=np(1-p)$.

We need to find n such that $P\left(\left|\overline{p}-p\right| \leq 0.01\right) \geq 0.95$

$$P\left(\left|\overline{p}-p\right| \leq 0.01\right) = 1 - P\left(\left|\overline{p}-p\right| > 0.01\right)$$

Because $P\left(\left|\overline{p}-p\right| > 0.01\right) \leq P\left(\left|\overline{p}-p\right| \geq 0.01\right)$, we have:

$$P\left(\left|\overline{p}-p\right| \leq 0.01\right) \geq 1 - P\left(\left|\overline{p}-p\right| \geq 0.01\right)$$

So as long as we can ensure that $1 - P\left(\left|\overline{p}-p\right| \geq 0.01\right) \geq 0.95$, we'll be fine. This gives us $P\left(\left|\overline{p}-p\right| \geq 0.01\right) \leq 0.05$

Using Chebyshev Inequality, we have:

$$P\left(\left|\overline{p}-E\left(\overline{p}\right)\right| \geq 0.01\right) \leq \frac{Var\left(\overline{p}\right)}{0.01^2}$$

$$\overline{p} = \frac{m}{n} \quad \Rightarrow \quad E\left(\overline{p}\right) = E\left(\frac{m}{n}\right) = \frac{E\left(m\right)}{n} = \frac{np}{n} = p,$$

$$Var\left(\overline{p}\right) = Var\left(\frac{m}{n}\right) = \frac{Var\left(m\right)}{n^2} = \frac{np\left(1-p\right)}{n^2} = \frac{p\left(1-p\right)}{n}$$

$$P\left(\left|\overline{p}-p\right| \geq 0.01\right) \leq \frac{Var\left(\overline{p}\right)}{0.01^2} = \frac{p\left(1-p\right)}{0.01^2 n}$$

So we need to ensure that $\dfrac{p\left(1-p\right)}{0.01^2 n} \leq 0.05$. This gives us $n \geq \dfrac{p\left(1-p\right)}{0.01^2\left(0.05\right)}$.

Are we stuck? No. We don't know $p\left(1-p\right)$, but we can find the upper bond of $p\left(1-p\right)$:

$$p\left(1-p\right) = p - p^2 = \frac{1}{4} - \left(p^2 - p - \frac{1}{4}\right) = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2 \leq \frac{1}{4}$$

So the max value of $\dfrac{p(1-p)}{0.01^2(0.05)}$ is $\dfrac{1}{4}\times\dfrac{1}{0.01^2(0.05)}=50,000$. If we let n take on this value, we can guarantee that $P\left(\left|\overline{p}-p\right|\geq 0.01\right)\leq 0.05$.

So if we survey 50,000 or more exam candidates and find the supporting ratio, chances are less than 5% that the sample supporting ratio will differ from the true supporting ratio by 0.01.

General formula:

We want to conduct a survey to estimate the percentage of people who does something. We want to make sure that the probability is at least b that the percentage found from the sample differs from the true percentage in the population by amount no more than a

Based Chebyshev Inequality, how many people do we need to survey?

Solution

Let p represent the true percentage of the people who do something. Let $\overline{p}$ represent the percentage found in surveyed.

Suppose we sample n people. m of the n people surveyed do something.

We are asked to find $P\left(\left|\overline{p}-p\right|\leq a\right)\geq b$.

$$\overline{p}=\frac{m}{n} \quad \Rightarrow \quad E\left(\overline{p}\right)=E\left(\frac{m}{n}\right)=\frac{E(m)}{n}=\frac{np}{n}=p,$$

$$Var\left(\overline{p}\right)=Var\left(\frac{m}{n}\right)=\frac{Var(m)}{n^2}=\frac{np(1-p)}{n^2}=\frac{p(1-p)}{n}\leq\frac{1}{4n}$$

Because $p=E\left(\overline{p}\right)$, $P\left(\left|\overline{p}-p\right|\leq a\right)\geq b$ is equivalent to $P\left(\left|\overline{p}-E\left(\overline{p}\right)\right|\leq a\right)\geq b$.

$$P\left(\left|\overline{p}-E\left(\overline{p}\right)\right|\leq a\right)=\left[1-P\left(\left|\overline{p}-E\left(\overline{p}\right)\right|>a\right)\right]>\left[1-P\left(\left|\overline{p}-E\left(\overline{p}\right)\right|\geq a\right)\right]$$

So if we make sure $\left[1-P\left(\left|\overline{p}-E\left(\overline{p}\right)\right|\geq a\right)\right]\geq b$, we'll have $P\left(\left|\overline{p}-p\right|\leq a\right)\geq b$.

Using Chebyshev Inequality, we have:

$$P\left(\left|\overline{p}-E\left(\overline{p}\right)\right|\geq a\right)\leq\frac{Var\left(\overline{p}\right)}{a^2}\leq\frac{1}{4na^2}$$

To ensure that $\left[1-P\left(\left|\overline{p}-E\left(\overline{p}\right)\right|\geq a\right)\right]\geq b$, we need to have:

$$1-\frac{1}{4na^2}\geq b,\ n\geq\frac{1}{4a^2\left(1-b\right)}$$

So we need to survey at least $\dfrac{1}{4a^2\left(1-b\right)}$ people.

Problem 4

Redo Problem 2 using normal approximation.

Solution

Just as before, we need to make sure

$$P\left(\left|\overline{p}-p\right|\geq 0.01\right)\leq 0.05$$

Next, we'll use normal approximation to calculate $P\left(\left|\overline{p}-p\right|\geq 0.01\right)$. We assume that the random variable $\overline{p}$ is normally distributed. As a result, $\overline{p}-E\left(p\right)$ is approximately normal.

$$E\left[\overline{p}-E\left(p\right)\right]=E\left(\overline{p}\right)-E\left(p\right)$$

Remember $E\left(p\right)$ is a constant.

From the last problem, we know that $E\left(\overline{p}\right)=E\left(p\right)$. So the normal random variable $\overline{p}-E\left(p\right)$ has a mean of zero.

The variance of $\overline{p}-E\left(p\right)$ is:

$$Var\left[\overline{p}-E\left(p\right)\right]=Var\left(\overline{p}\right)-Var\left[E\left(p\right)\right]=Var\left(\overline{p}\right)$$

Remember $E\left(p\right)$ is a constant. So its variance is zero.

From the last problem, we know

$$Var\left(\overline{p}\right)=Var\left(\frac{m}{n}\right)=\frac{p\left(1-p\right)}{n}\leq\frac{1}{4n}$$

So we'll use $\frac{1}{4n}$ as the variance of $Var\left(\overline{p}\right)$. This is the worse case scenario. The n calculated based on this largest variance is safe.

Because $\overline{p}-E\left(p\right)$ is approximately normal, we have

$$P\left(\left|\overline{p}-E\left(p\right)\right|\geq 0.01\right)=2P\left\{\left[\overline{p}-E\left(p\right)\right]\geq 0.01\right\}$$

$$\Rightarrow\quad P\left(\left|\overline{p}-p\right|\geq 0.01\right)=P\left(\left|\overline{p}-E\left(p\right)\right|\geq 0.01\right)=2P\left\{\left[\overline{p}-E\left(p\right)\right]\geq 0.01\right\}$$

$$P\left\{\left[\overline{p}-E\left(p\right)\right]\geq 0.01\right\}=1-P\left\{\left[\overline{p}-E\left(p\right)\right]\leq 0.01\right\}$$

$$=1-\Phi\left\{\frac{0.01-E\left[\overline{p}-E\left(p\right)\right]}{\sqrt{Var\left[\overline{p}-E\left(p\right)\right]}}\right\}=1-\Phi\left[\frac{0.01}{\sqrt{\frac{1}{4n}}}\right]=1-\Phi\left(0.02\sqrt{n}\right)$$

$$P\left(\left|\overline{p}-p\right|\geq 0.01\right)=2P\left\{\left[\overline{p}-E\left(p\right)\right]\geq 0.01\right\}=2\left[1-\Phi\left(0.02\sqrt{n}\right)\right]$$

$$P\left(\left|\overline{p}-p\right|\geq 0.01\right)\leq 0.05\quad\Rightarrow\quad 2\left[1-\Phi\left(0.02\sqrt{n}\right)\right]\leq 0.05,\ \Phi\left(0.02\sqrt{n}\right)\geq 0.975$$

$$\Rightarrow\quad 0.02\sqrt{n}\geq 41.96,\ n\geq 9,604$$

We see that normal approximation gives us a far smaller n than does Chebyshev Inequality.

Problem 5

A poll is conducted among Exam P candidates to estimate the proportion of Exam P candidates who use "Deeper Understanding, Faster Calc" (the Guo manual). We want to make sure that the probability is at least 80% that the sample proportion will not differ from the true proportion by at most 0.1.

How many people do we need to survey? Use Chebyshev Inequality.

Solution

We want $P\left(\left|\overline{p}-p\right|\leq 0.1\right)\geq 0.8$. Here $a=0.1$ and $b=0.8$.

$$n\geq\frac{1}{4a^2(1-b)}=\frac{1}{4\left(0.1^2\right)(1-0.8)}=125$$

Problem 6

Let X represent the hours it takes a randomly chosen Exam P candidate to prepare for Exam P. Suppose we know, from other information, that the variance of X is no more than 40, i.e. $Var(X)\leq 40$.

How large is the sample required to make sure that the probability is at least 0.95 that the sample mean differs from the true mean $E(X)$ by no more than 0.5?

Solution

Here we can't use the formula $n\geq\frac{1}{4a^2(1-b)}$. Here we are interested in finding the average # of hours it takes an Exam P candidate to prepare for Exam P. The # of hours is not a binomial distribution. So we can't use the formula $n\geq\frac{1}{4a^2(1-b)}$.

The solution is still simple. Assume we sample n people. Then,

$$\overline{X}=\frac{X_1+X_2+...+X_n}{n}$$

Here $X_1, X_2, ..., X_n$ are independent identically distributed.

$$\Rightarrow\quad E\left(\overline{X}\right)=E\left(\frac{X_1+X_2+...+X_n}{n}\right)=\frac{nE(X)}{n}=E(X)$$

$$\Rightarrow\quad Var\left(\overline{X}\right)=Var\left(\frac{X_1+X_2+...+X_n}{n}\right)=\frac{nVar(X)}{n^2}=\frac{Var(X)}{n}$$

We need to make sure $P\left(\left|\overline{X}-E(X)\right|\leq 0.5\right)\geq 0.95$. Because $E\left(\overline{X}\right)=E(X)$, we need to ensure that

$$P\left(\left|\overline{X}-E\left(\overline{X}\right)\right|\leq 0.5\right)\geq 0.95$$

$$P\left(\left|\overline{X}-E\left(\overline{X}\right)\right|\leq 0.5\right) \quad \geq \quad \left[1-P\left(\left|\overline{X}-E\left(\overline{X}\right)\right|\geq 0.5\right)\right]$$
$$\Rightarrow \quad 1-P\left(\left|\overline{X}-E\left(\overline{X}\right)\right|\geq 0.5\right) \quad \geq \quad 0.95$$
$$\Rightarrow \quad P\left(\left|\overline{X}-E\left(\overline{X}\right)\right|\geq 0.5\right) \quad \leq \quad 0.05$$

Using Chebyshev Inequality, we have:

$$P\left(\left|\overline{X}-E\left(\overline{X}\right)\right|\geq 0.5\right) \quad \leq \quad \frac{Var\left(\overline{X}\right)}{0.5^2}=\frac{Var\left(X\right)}{0.5^2 n}\leq\frac{40}{0.5^2 n}$$

We need to make sure $1-\dfrac{Var\left(X\right)}{0.5^2 n}\geq 0.95$

$$1-\frac{40}{0.5^2 n}\geq 0.95, \qquad \Rightarrow \quad n\geq 3,200$$

In general, to ensure $P\left(\left|\overline{X}-E\left(\overline{X}\right)\right|\leq a\right)\geq b$, we need to survey at least $n\geq\dfrac{Var\left(X\right)}{a^2\left(1-b\right)}$ people based on Chebyshev Inequality.

Problem 7

We know that $V\left(X\right)\leq 1$. What's the sample size to ensure that $P\left(\left|\overline{X}-E\left(\overline{X}\right)\right|\leq 0.2\right)\geq 0.95$?

Solution

$$n\geq\frac{Var\left(X\right)}{a^2\left(1-b\right)}\leq\frac{1}{0.2^2\left(1-0.95\right)}=500$$

So we need to have a sample size of at least 500.

Chapter 36 Study Note "Risk and Insurance" explained

This chapter explains the study note titled "Risk and Insurance" by Anderson and Brown.

Deductible, benefit limit

An insurance contract has a deductible d per loss and maximum payment (called benefit limit) u per loss. Let X represent the loss incurred by the policyholder. Let Y represent the payment made the insurer to the policyholder. Then

$$Y = \begin{cases} 0 & \text{if} \quad 0 \le X \le d \\ X - d & \text{if} \quad d \le X \le d + u \\ u & \text{if} \quad X \ge d + u \end{cases}$$

Example 1

You bought an insurance policy. The deductible is \$200 per loss. The insurance company will pay the maximum of \$5,000 per loss. One day you suffered a loss of \$150. How much will the insurer pay you? How much do you have to pay out of your own pocket to cover your loss?

Solution

Your loss is less than the deductible. As a result, the insurance benefit won't kick in. You need to pay \$150 to cover the loss. The insurance company pays you none.

Example 2

You bought an insurance policy. The deductible is \$200 per loss. The insurance company will pay the maximum of \$5,000 per loss. One day you suffered a loss of \$600. How much will the insurer pay you? How much do you have to pay out of your own pocket to cover your loss?

Solution

You will pay the 1^{st} \$200. After you have met the deductible, the insurance benefits will kick in. The insurance company will pay you \$600-\$200=\$400.

Example 3

You bought an insurance policy. The deductible is \$200 per loss. The insurance company will pay a maximum of \$5,000 per loss. One day you suffered a loss of \$6,000. How much will the insurer pay you? How much do you have to pay out of your own pocket to cover your loss?

Solution

You will pay the 1st $200 and then the insurance benefit will kick in. If there is no cap on the how much the insurance company will pay you per loss incident, you'll get $6,000-$200=$5,800 from the insurer. However, the benefits payment is capped at $5,000 per loss. As a result, the insurance company will pay you only $5,000, not $5,800. You'll need to cover the remaining loss $5,800-$5,000=$800 with your own money. Your total out-of-pocket cost is $200+$800=$1,000.

Example 4

A car owner has a 70% chance of no accidents in a year, 30% chance of having only one accident in a year, and 0% chance of having more than one accident in a year.

If there's an accident, the loss amount (also called "severity") is a random variable with the following distribution:

Severity	Probability
20	0.10
60	0.25
150	0.30
200	0.35

There's an annual deductible of $50. The annual maximum payment by the insurer is $145.

Calculate
(1) Annual expected loss incurred by a car owner
(2) Standard deviation of the annual loss incurred by the car owner
(3) Annual expected payment made by the insurance company to a car owner
(4) Standard deviation of the payment made by the insurance company to a car owner
(5) Annual expected cost that the insured car owner must cover his loss with his own money
(6) Standard deviation of the cost that the insured car owner must cover his loss with his own money
(7) Correlation coefficient between the insurance company's annual payment and the insured's annual out-of-pocket cost to cover the loss.

Solution

Questions (1) and (2)

Let X represent the annual loss incurred by a car owner.

$$f(x)=\begin{cases} \quad\quad\quad\quad\quad 70.0\% & x=0 \\ 30\%(0.10)=\ 3.0\% & x=20 \\ 30\%(0.25)=\ 7.5\% & x=60 \\ 30\%(0.30)=\ 9.0\% & x=150 \\ 30\%(0.35)=10.5\% & x=200 \end{cases}$$

In problems like this one, always thoroughly list all of the possibilities. In addition, make sure the total probabilities add up to one. Here 70%+3%+7.5%+9%+10.5%=100%. Good.

$$E(X)=\sum xf(x)=0(70\%)+20(3\%)+60(7.5\%)+150(9\%)+200(10.5\%)=39.6$$

$$\begin{aligned} Var(X)&=\sum\left[x-E(x)\right]^2 f(x) \\ &=(0-39.6)^2(70\%)+(20-39.6)^2(3\%)+(60-39.6)^2(7.5\%) \\ &\quad+(150-39.6)^2(9\%)+(200-39.6)^2(10.5\%) \\ &=4{,}938.84 \end{aligned}$$

$$\sigma=\sqrt{Var(X)}=70.28$$

To avoid the above calculations, scale $f(x)$ to an integer and enter the data pairs of $\left[x,\ f(x)\right]$ to BA II Plus 1-V Statistics Worksheet.

Questions (3) and (4)

Let Y represent the annual payment made by the insurer to the insured. Then

$$f(y)=\begin{cases} \quad\quad\quad\quad\quad 70.0\% & y=0 & \text{when } x=0 \\ 30\%(0.10)=\ 3.0\% & y=0 & \text{when } x=20 \\ 30\%(0.25)=\ 7.5\% & y=60-50=10 & \text{when } x=60 \\ 30\%(0.30)=\ 9.0\% & y=150-50=100 & \text{when } x=150 \\ 30\%(0.35)=10.5\% & y=145 & \text{when } x=200 \end{cases}$$

When the loss $X=200$, the insured will pay \$150 if there's no benefit limit. However, since there's a limit of \$145, the insurer's payment is capped at \$145.

The distribution of Y can be simplified as follows:

y	0	10	100	145
$f(y)$	73%	7.50%	9%	10.50%

Using BA II Plus or memorized formulas, we have:

$$E(Y)=24.975,\quad \sigma_Y=49.91$$

Questions (5) and (6)

Let Z represent how much money the insured needs take out from his own pocket to cover his annual loss.

$$f(z)=\begin{cases} \quad\quad\quad\quad\quad 70.0\% & z=0 \quad \text{when } x=0 \\ 30\%(0.10)=3.0\% & z=20 \quad \text{when } x=20 \\ 30\%(0.25)=7.5\% & z=50 \quad \text{when } x=60 \\ 30\%(0.30)=9.0\% & z=50 \quad \text{when } x=150 \\ 30\%(0.35)=10.5\% & z=55 \quad \text{when } x=200 \end{cases}$$

Please note that when $x=200$, the insurer will pay only \$145. As a result, the insured needs to pay the remainder of \$55 to cover his loss.

The distribution of Z can be simplified as follows:

z	0	20	50	55
$f(z)$	70%	3%	16.5%	10.50%

Using BA II Plus or memorized formulas, we have:

$$E(Z)=14.625,\quad \sigma_Z=22.98$$

Question (7)

We are asked to calculate $\rho_{Y,Z}=\dfrac{Cov(Y,Z)}{\sigma_Y\sigma_Z}$. You need to remember the core equation:

$$\underbrace{X}_{\text{loss}}=\underbrace{Y}_{\text{insurer's share}}+\underbrace{Z}_{\text{insured's share}}$$

The above equation says that if there's a loss, either the insurer pays or the insured pays or both. As a result, the loss amount must be equal to the sum of what the insurer will pay and what the insured will pay. We can verify that $X=Y+Z$ holds under every scenario:

$$f(x)=\begin{cases} \quad 70.0\% & x=0 \\ 30\%(0.10)=\ 3.0\% & x=20 \\ 30\%(0.25)=\ 7.5\% & x=60 \\ 30\%(0.30)=\ 9.0\% & x=150 \\ 30\%(0.35)=10.5\% & x=200 \end{cases}$$

$$f(y)=\begin{cases} \quad 70.0\% & y=0 & \text{when } x=0 \\ 30\%(0.10)=\ 3.0\% & y=0 & \text{when } x=20 \\ 30\%(0.25)=\ 7.5\% & y=60-50=10 & \text{when } x=60 \\ 30\%(0.30)=\ 9.0\% & y=150-50=100 & \text{when } x=150 \\ 30\%(0.35)=10.5\% & y=145 & \text{when } x=200 \end{cases}$$

$$f(z)=\begin{cases} \quad 70.0\% & z=0 & \text{when } x=0 \\ 30\%(0.10)=\ 3.0\% & z=20 & \text{when } x=20 \\ 30\%(0.25)=\ 7.5\% & z=50 & \text{when } x=60 \\ 30\%(0.30)=\ 9.0\% & z=50 & \text{when } x=150 \\ 30\%(0.35)=10.5\% & z=55 & \text{when } x=200 \end{cases}$$

$$\Rightarrow \quad E(X)=E(Y)+E(Z)$$

As a check, we know $E(X)=39.6$, $E(Y)=24.975$, and $E(Z)=14.625$. 39.6=24.975+14.625. The equation holds. Now we know that our calculations of $E(X)$, $E(Y)$, and $E(Z)$ are OK.

$$\Rightarrow \quad Var(X)=Var(Y+Z)=Var(Y)+Var(Z)+2Cov(Y,Z)$$

$$\Rightarrow \quad 2Cov(Y,Z)=Var(X)-\left[Var(Y)+Var(Z)\right]=70.28^2-\left(49.91^2+22.98^2\right)=1{,}920.19$$

$$\Rightarrow \quad Cov(Y,Z)=\frac{1{,}920.19}{2}=960.095$$

$$\Rightarrow \quad \rho_{Y,Z}=\frac{Cov(Y,Z)}{\sigma_Y\sigma_Z}=\frac{960.095}{49.91(22.98)}=0.837$$

We shouldn't be surprised that $\rho_{Y,Z}$ is close to 1; if you get $\rho_{Y,Z}$ close to zero, your calculation must be wrong. Y and Z should have a good linear relationship.

Coinsurance

Some insurance contracts have a coinsurance factor. An insurance policy has a deductible of d and a benefit limit (the insurer's max payment) of u. The insurer's portion of the payment is (coinsurance factor) is α where $0<\alpha\le 1$. Let X represent the loss and Y represent the insurer's payment. Then

$$Y=\begin{cases} 0 & \text{if} \quad 0\le X\le d \\ \alpha\left(X-d\right) & \text{if} \quad d\le X\le d+\dfrac{u}{\alpha} \\ u & \text{if} \quad X\ge d+\dfrac{u}{\alpha} \end{cases}$$

Example 5

You bought an insurance policy. The deductible is $200 per loss. The insurance company will pay 90% of the loss in excess of the deductible subject to the maximum payment of $5,000 per loss. One day you suffered a loss of $4,000. How much will the insurer pay you? How much do you have to pay out of your own pocket to cover your loss?

Solution

The insurer will pay 90%(4,000-200)=$3,420. You will need to cover the remaining $4,000 - 3,420 = $580 loss out of your own money.

Example 6

You bought an insurance policy. The deductible is $200 per loss. The insurance company will pay 90% of the loss in excess of the deductible subject to the maximum payment of $5,000 per loss. One day you suffered a loss of $5,750. How much will the insurer pay you? How much do you have to pay out of your own pocket to cover your loss?

Solution

The insurer will pay you 90%(5,750-200)=$4,995. You'll need to cover the remaining $5,750 – 4,995 = $755 out of your own pocket.

Example 7

You bought an insurance policy. The deductible is $200 per loss. The insurance company will pay 90% of the loss in excess of the deductible subject to the maximum payment of $5,000 per loss. One day you suffered a loss of $5,760. How much will the insurer pay you? How much do you have to pay out of your own pocket to cover your loss?

Solution

If there's no benefit limit, then the insurer will pay you 90%(5,760-200)=$5,004. However, the insurer's payment is capped at $5,000. So the insurer will pay you $5,000. You'll need to cover the remaining $5,760 – 5,000 = $760 out of your own pocket.

Example 8

A car owner has 70% chance of no accidents in a year, 30% chance of having only one accident in a year, and 0% chance of having more than one accident in a year.

If there's an accident, the loss amount (also called "severity") is a random variable with the following distribution:

Severity	Probability
20	0.10
60	0.25
150	0.30
200	0.35

There's an annual deductible of $50. The annual maximum payment by the insurer is $85. The insurer will pay 60% of the loss in excess of the deductible subject to the max annual payment.

Calculate

(1) Annual expected loss incurred by a car owner
(2) Standard deviation of the annual loss incurred by the car owner
(3) Annual expected payment made by the insurance company to a car owner
(4) Standard deviation of the payment made by the insurance company to a car owner
(5) Annual expected cost that the insured car owner must cover his loss with his own money
(6) Standard deviation of the cost that the insured car owner must cover his loss with his own money
(7) Correlation coefficient between the insurance company's annual payment and the insured's annual out-of-pocket cost to cover the loss.

Solution

Questions (1) and (2)

Let X represent the annual loss incurred by a car owner.

$$f(x)=\begin{cases} \quad\quad\quad\quad 70.0\% & x=0 \\ 30\%(0.10)=\ 3.0\% & x=20 \\ 30\%(0.25)=\ 7.5\% & x=60 \\ 30\%(0.30)=\ 9.0\% & x=150 \\ 30\%(0.35)=10.5\% & x=200 \end{cases}$$

$$E(X)=\sum xf(x)=0(70\%)+20(3\%)+60(7.5\%)+150(9\%)+200(10.5\%)=39.6$$

$$\begin{aligned} Var(X) &= \sum\left[x-E(x)\right]^2 f(x) \\ &= (0-39.6)^2(70\%)+(20-39.6)^2(3\%)+(60-39.6)^2(7.5\%) \\ &\quad +(150-39.6)^2(9\%)+(200-39.6)^2(10.5\%) \\ &\Rightarrow Var(X)=4{,}938.84,\ \sigma=\sqrt{Var(X)}=70.28 \end{aligned}$$

Once again, you can use BA II Plus 1-V Statistics Worksheet to do the calculations.

Questions (3) and (4)

Let Y represent the annual payment made by the insurer to the insured. Then

$$f(y)=\begin{cases} \quad\quad\quad\quad 70.0\% & y=0 & \text{when } x=0 \\ 30\%(0.10)=\ 3.0\% & y=0 & \text{when } x=20 \\ 30\%(0.25)=\ 7.5\% & y=0.6(60-50)=6 & \text{when } x=60 \\ 30\%(0.30)=\ 9.0\% & y=0.6(150-50)=60 & \text{when } x=150 \\ 30\%(0.35)=10.5\% & y=85 & \text{when } x=200 \end{cases}$$

When the loss $X=200$, the insured will pay 0.6(200-50)=90 if there's no benefit limit. However, since there's a limit of \$85, the insurer's payment is capped at \$85.

The distribution of Y can be simplified as follows:

y	0	6	60	85
$f(y)$	73%	7.50%	9%	10.50%

Using BA II Plus or memorized formulas, we have:

$E(Y)=14.775$, $\sigma_Y=29.45$

Questions (5) and (6)

Let Z represent how much money the insured needs take out from his own pocket to cover his annual loss.

$$f(z)=\begin{cases} 70.0\% & z=0 & \text{when } x=0 \\ 30\%(0.10)=\ 3.0\% & z=20 & \text{when } x=20 \\ 30\%(0.25)=\ 7.5\% & z=50+0.4(60-50)=54 & \text{when } x=60 \\ 30\%(0.30)=\ 9.0\% & z=50+0.4(150-50)=90 & \text{when } x=150 \\ 30\%(0.35)=10.5\% & z=200-85=115 & \text{when } x=200 \end{cases}$$

The distribution of Z can be simplified as follows:

z	0	20	54	90	115
$f(z)$	70%	3%	7.5%	9%	10.5%

Using BA II Plus or memorized formulas, we have:

$E(Z)=24.825$, $\sigma_Z=41.62$

Check: $E(Y)+E(Z)=14.775+24.825=39.6=E(X)$. OK.

Question (7)

$\Rightarrow \quad Var(X)=Var(Y+Z)=Var(Y)+Var(Z)+2Cov(Y,Z)$

$\Rightarrow \quad 2Cov(Y,Z)=Var(X)-[Var(Y)+Var(Z)]=70.28^2-(29.45^2+41.62^2)=2,339.75$

$\Rightarrow \quad Cov(Y,Z)=\dfrac{2,339.75}{2}=1,169.88$

$\Rightarrow \quad \rho_{Y,Z}=\dfrac{Cov(Y,Z)}{\sigma_Y\sigma_Z}=\dfrac{1,169.88}{29.45(41.62)}=0.95$

The effect of inflation on loss and claim payment.

This topic is covered in depths in Exam M and C. For Exam P, you just need to learn the basics. If you understand the examples in the study note, you should be fine.

Problem 9

Looking again at the 100 insured car owners with a 500 deductible and no benefit limit, assume that there's 10% annual inflation. Over the next 5 years, what would the expected claim payments and the insurer's risk be?

The study note gives you the following table. Reproduce this table.

	$f(y,t)$	80%	10%	8%	2%	Expected claim payment	Standard deviation of claim payments
Year							
1	Loss	\$0	\$500	\$5,000	\$15,000	\$750	\$2,324
1	Claim	\$0	\$0	\$4,500	\$14,500	\$650	
2	Loss	\$0	\$550	\$5,500	\$16,500	\$825	\$2,568
2	Claim	\$0	\$50	\$5,000	\$16,000	\$725	
3	Loss	\$0	\$605	\$6,050	\$18,150	\$908	\$2,836
3	Claim	\$0	\$105	\$5,550	\$17,650	\$808	
4	Loss	\$0	\$666	\$6,655	\$19,965	\$998	\$3,131
4	Claim	\$0	\$166	\$6,155	\$19,465	\$898	
5	Loss	\$0	\$732	\$7,321	\$21,962	\$1,098	\$3,456
5	Claim	\$0	\$232	\$6,821	\$21,462	\$998	

Solution

Year 1

This is the base year so inflation has no effect. The loss amounts (severities) are \$0, \$500, \$5,000, and \$15,000. Since there's a deductible of \$500, the insurance company will pay \$0, \$500-500=\$0, \$5,000-500=\$4,500, and \$15,000-500=\$14,500 with probabilities of 80%, 10%, 8%, and 2% respectively.

Expected loss:

0(80%)+500(10%)+5,000(8%)+15,000(2%)=\$750

Expected claim payments by the insurer:

0(80%)+0(10%)+4,500(8%)+14,500(2%)=\$650

Standard deviation of claims:

$$Var=(0-650)^2\,0.8+(0-650)^2\,0.1+(4,500-650)^2\,0.08+(14,500-650)^2\,0.02$$

$\Rightarrow$ Var =5,402,500, $\sigma=\sqrt{Var}=2,324.33$

Year 2:

Due to inflation, losses have increased 10%. The loss amounts are:

\$0(1.1)=\$0, \$500(1.1)=\$550, \$5,000(1.1)=\$5,500, \$15,000(1.1)=\$16,500

The claims after the deductible are:

\$0, \$550-500=\$50, \$5,500 – 500=\$5,000, \$16,500-500=\$16,000.

Expected loss:

0(80%)+550(10%)+5,500(8%)+16,500(2%)=\$825

Expected claim:

0(80%)+50(10%)+5,000(8%)+16,000(2%)=\$725

Standard deviation of the claims:

$$\sqrt{(0-725)^2\,0.8+(50-725)^2\,0.1+(5,000-725)^2\,0.08+(16,000-725)^2\,0.02}=2,568$$

Year 3:

Due to inflation, losses have increased another 10%. The loss amounts are:

\$0(1.1^2)=\$0; \$500(1.1^2)=\$605;
\$5,000(1.1^2)=\$6,050; \$15,000(1.1^2)=\$18,150

The claims after the deductible are:
\$0, \$605-500=\$105, \$6,050 – 500=\$5,550, \$18,150-500=\$17,650.
Plugging into the mean and variance formulas, you should get:

Expected loss = \$908, Expected claim=\$808

Standard deviation of the claim = \$2,836

Year 4 and 5:

Just increase the losses by 10% each year. You should be able to reproduce the results (there may be rounding differences).

Observation

Please note that the expected loss also increases by 10% each year, yet the expected claim and the standard deviation of the claim increased by more than 10%. Make sure you understand why.

Let's look at the expected loss:

Year	1	2	3	4	5
Expected loss	\$750	\$825	\$908	\$998	\$1,098

825=750(1.1); 908=825(1.1) = 750(1.1^2);
998=908(1.1)= 750(1.1^3); 1,098=998(1.1)= 750(1.1^4).

Why does the expected loss year after year increase by 10% each year? If losses x_i increase by uniform percentage, then $E(X)=\sum x_i f(x_i)$ should increase by the same percentage.

Why do the expected claims year after year increase by more than 10%? Let Y represent the claim payment. The formula is:

$$Y = \min(X-d,0)$$

$$E(Y)=\sum \min(x_i-d,0)\, f(x_i)$$

Now x_i increase by uniform percentage each year, but the deductible d is fixed. Consequently, $E(Y)$ increases by more than 10% each year. Similarly, σ_Y also increases by 10% each year because we have a fixed deductible.

Problem 10

Looking again at the 100 insured car owners with a 500 deductible and benefit limit of 12,500, assume that there's 10% annual inflation. Over the next 5 years, what would the expected claim payments and the insurer's risk be?

The study note tells us that at the end of Year 5, the expected loss is 1,098; the expected claim is $819; and the standard deviation of the claim is 2,486.

Reproduce the results for Year 5.

Solution

Probability	80%	10%	8%	2%
Loss at end of Year 1	$0	$500	$5,000	$15,000
Loss at end of Year 5	$0	$732.05 =500(1.1^4)	$7,320.5 =5,000(1.1^4)	$21,961.5 =15,000(1.1^4)
Claim at end of Year 5	$0	$232.05 =732.05-500	$6,820.5 =$7,320.50-500	$12,500 =min(21,961.5-500, 12,500)

If we let X =loss at the end of Year 5 and Y =claim at the end of Year 5, the above table can be simplified as follows:

Probability	80%	10%	8%	2%
X	$0	$732.05	$7,320.5	$21,961.5
Y	$0	$232.05	$6,820.5	$1,2,500

Using BA II Plus 1-V Statistics Worksheet, we have:

$E(X)=1,098.075$, $E(Y)=818.845$, $\sigma_Y=2,486.245$

Mixture of distributions

Section VIII of the study note touches up a very important concept "mixture of distributions." Since SOA can easily come up a similar question in the exam, let's go through this concept. I'll walk you through the examples used in the study note.

Problem 11

Consider an insurance policy that reimburses annual hospital charges for an insured individual. The probability of any individual being hospitalized in a year is 15%. That is,

$P(H=1)=0.15$. Once an individual is hospitalized, the charges X have a probability density function (pdf) $f_X(x|H=1)=0.1e^{-0.1x}$ for $x>0$.

Determine the expected value, the standard deviation, and the ratio of the standard deviation to the mean (coefficient of variation) of hospital charges for an insured individual.

Solution

I'll provide an alternative, simpler solution. Let Y represent hospital charges. Then

$$Y=\begin{cases} 0 & \text{with probability of 85\%} \\ X & \text{with probability of 15\%, where } X \text{ is distributed as } f(x)=0.1\,e^{-0.1x} \end{cases}$$

Then

$$E(Y)=85\%E(0)+15\%E(X)=15\%E(X)$$
$$E(Y^2)=85\%E(0^2)+15\%E(X^2)=15\%E(X^2)$$
$$Var(Y)=E(Y^2)-E^2(Y)=0.15E(X^2)-\left[0.15E(X)\right]^2=0.15E(X^2)-0.15^2E^2(X)$$

Here we use the following formula: if k is a constant, then $E(k)=k$ and $E(k^2)=k^2$. To understand this formula, imagine you have a random variable X whose density is $f_X(x)=1$ if $x=k$ and zero otherwise. Then $E(X)=k$ and $E(k^2)=k^2$.

Let's continue. X is an exponential random variable with mean of 10.

$$\Rightarrow \quad E(X)=\sigma_X=10,\ E(X^2)=E^2(X)+\sigma_X^2=200$$

$$\Rightarrow \quad E(Y)=15\%E(X)=15\%(10)=1.5$$

$$\Rightarrow \quad Var(Y)=0.15E(X^2)-0.15^2E^2(X)=0.15(200)-0.15^2(10^2)=27.75$$

The coefficient of variation is $\dfrac{\sigma_Y}{E(Y)}=\dfrac{\sqrt{27.75}}{1.5}=3.51$

In this example, Y is a mixture of random variables 0 and X with the following distribution:

$$Y = \begin{cases} 0 & \text{with probability of } 85\% \\ X & \text{with probability of } 15\%, \text{ where } X \text{ is distributed as } f(x) = 0.1e^{-0.1x} \end{cases}$$

One mistake made by many is to write:
$Y = 0.85(0) + 0.15X = 0.15X$ (Wrong!)

If you write $Y = 0.85(0) + 0.15X = 0.15X$, then you are treating Y as a fraction of X. Then Y is no longer a mixture of two random variables.

The wrong expression $Y = 0.85(0) + 0.15X = 0.15X$ leads the following wrong result:

$E(Y^2) = E(0.15X)^2 = 0.15^2 E(X^2)$ (Wrong!)
The correct formula is $E(Y^2) = 85\%E(0^2) + 15\%E(X^2) = 0.15E(X^2)$

$Var(Y) = Var(0.15X) = 0.15^2 Var(X) = 0.15^2 E(X^2) - 0.15^2 E^2(X)$ (Wrong!)
The correct formula is:
$Var(Y) = E(Y^2) - E^2(Y) = 0.15E(X^2) - [0.15E(X)]^2 = 0.15E(X^2) - 0.15^2 E^2(X)$

This brings up a critical point: the difference between a mixture and a sum. Say two random variables X and Y. X is distributed as follows:

$$X = \begin{cases} W_1 & \text{with probability of } p \\ W_2 & \text{with probability of } 1-p \end{cases}$$

Then you are given $Y = pW_1 + (1-p)W_2$.

Is $X = Y$? No. Here X is a mixture and Y is the sum. As said earlier, you can't write: $X = pW_1 + (1-p)W_2$.

To see the difference between X and Y, notice:

$$E(X) = pE(W_1) + (1-p)E(W_2)$$
$$E(X^2) = pE(W_1^2) + (1-p)E(W_2^2)$$

$$E(Y) = E[pW_1 + (1-p)W_2] = pE(W_1) + (1-p)E(W_2)$$
$$E(Y^2) = E[pW_1 + (1-p)W_2]^2 = E[p^2W_1^2 + 2p(1-p)W_1W_2 + (1-p)^2 W_2^2]$$

Though $E(X)=E(Y)$, clearly $E(X^2)\neq E(Y^2)$.

Problem 12

Consider an insurance policy that reimburses annual hospital charges for an insured individual. The probability of any individual being hospitalized in a year is 15%. That is, $P(H=1)=0.15$. Once an individual is hospitalized, the charges X have a probability density function (pdf) $f_X(x|H=1)=0.1\,e^{-0.1x}$ for $x>0$.

Assume there's a deductible of 5

Determine the expected value, the standard deviation, and the ratio of the standard deviation to the mean (coefficient of variation) of the claim payment.

Alternative solution

Let Y represent claims for hospital charges. Then

$$Y=\begin{cases}0 & \text{with probability of 85\%}\\ Z=\max(X-5,0) & \text{with probability of 15\%, where } X \text{ is distributed as } f(x)=0.1e^{-0.1x}\end{cases}$$

$$E(Y)=85\%E(0)+15\%E(Z)=15\%E(Z)$$
$$E(Y^2)=85\%E(0^2)+15\%E(Z^2)=15\%E(Z^2)$$
$$Var(Y)=E(Y^2)-E^2(Y)$$

$$E(Z)=\int_0^{+\infty}\max(x-5,0)\,f(x)\,dx=\int_0^{5}\max(x-5,0)\,f(x)\,dx+\int_5^{+\infty}\max(x-5,0)\,f(x)\,dx$$

If $x\leq 5$, then $\max(x-5,0)=0$. In other words, if a charge is \$5 or less, the insurer won't pay anything.

If $x\geq 5$, then $\max(x-5,0)=x-5$. The insurer pays the charge above and beyond the deductible of \$5.

$$\Rightarrow \quad E(Z)=\int_5^{+\infty}(x-5)\,f(x)\,dx=\int_5^{+\infty}(x-5)\frac{1}{10}e^{-\frac{x}{10}}dx$$

Set $u=x-5$.

$$\Rightarrow \quad E(Z)=\int_5^{+\infty}(x-5)f(x)dx=\int_0^{+\infty}u\frac{1}{10}e^{-\frac{u+5}{10}}du=e^{-0.5}\int_0^{+\infty}u\left(\frac{1}{10}e^{-\frac{u}{10}}\right)du=10e^{-0.5}$$

$$\Rightarrow \quad E(Y)=15\%E(Z)=15\%E(Z)=15\%\left(10e^{-0.5}\right)=1.5e^{-0.5}=0.91$$

Similarly,

$$E\left(Z^2\right)=\int_0^{+\infty}z^2 f(x)dx=\int_0^5\left[\max(x-5,0)\right]^2 f(x)dx+\int_5^{+\infty}\left[\max(x-5,0)\right]^2 f(x)dx$$

$$=\int_5^{+\infty}(x-5)^2 f(x)dx=\int_5^{+\infty}(x-5)^2\frac{1}{10}e^{-\frac{x}{10}}dx=\int_0^{+\infty}u^2\frac{1}{10}e^{-\frac{u+5}{10}}du$$

$$=e^{-0.5}\int_0^{+\infty}u^2\frac{1}{10}e^{-\frac{u}{10}}du=200e^{-0.5}$$

$$E\left(Y^2\right)=15\%E\left(Z^2\right)=15\%\left(200e^{-0.5}\right)=30e^{-0.5}$$

$$Var(Y)=E\left(Y^2\right)-E^2(Y)=30e^{-0.5}-\left(1.5e^{-0.5}\right)^2=17.3682$$

$$\sigma_Y=\sqrt{Var(Y)}=4.17$$

The coefficient of variation is: $\dfrac{\sigma_Y}{E(Y)}=\dfrac{4.17}{0.91}=4.58$

Coefficient of variation

Problem 13

You are given information about a block of insurance policies:

Class	# of policies	Probability that each policy has a loss in a year	distribution of the annual loss per policy if there's a loss	
			Mean	Variance
1	100	0.1	1	3
2	200	0.3	2	5

Assume all these 300 policies are independent.

Let S represent the total annual loss incurred by the 300 policies.

Calculate $\dfrac{\sigma_S}{E(S)}$, the coefficient of variation of S.

Solution

This problem looks scary. The key to solving this problem is to thoroughly list all of the random variables.

Let X_i represent the annual loss incurred by the i-th policy in Class 1.
Let Y_j represent the annual loss incurred by the j-th policy in Class 2.

We have $i=1,2,...,100$ and $j=1,2,...,200$

Then $S = X_1 + X_2 + ... + X_{100} + Y_1 + Y_2 + ... + Y_{200}$.

$$\Rightarrow \quad E(S) = E(X_1 + X_2 + ... + X_{100} + Y_1 + Y_2 + ... + Y_{200})$$

X_i's are independent identically distributed,

$$\Rightarrow \quad E(X_1) = E(X_2) = ... = E(X_{100}) = E(X)$$
$$\Rightarrow \quad Var(X_1) = Var(X_2) = ... = Var(X_{100}) = Var(X)$$

Here X refer to the annual loss incurred by a policy randomly chosen from Class 1.

Similarly, Y_i's are independent identically distributed,

$$\Rightarrow \quad E(Y_1) = E(Y_2) = ... = E(Y_{100}) = E(Y)$$
$$\Rightarrow \quad Var(Y_1) = Var(Y_2) = ... = Var(Y_{200}) = Var(Y)$$

Here Y refer to the annual loss incurred by a policy randomly chosen from Class 2.

$$\Rightarrow \quad E(S) = E(X_1) + E(X_2) + ... + E(X_{100}) + E(Y_1) + E(Y_2) + ... + E(Y_{200})$$
$$\Rightarrow \quad E(S) = 100E(X) + 200E(Y)$$
$$\Rightarrow \quad Var(S) = 100Var(X) + 200Var(Y)$$

Please note that one common mistake is to write:

$$S = 100X + 200Y \quad \text{(Wrong)}$$

Though X_1, X_2, …, X_{100} are independent identically distributed having the same mean $E(X)$ and the same variance $Var(X)$, X_1, X_2, …, X_{100} are not necessarily identical. In other words, we have $E(X_1)=E(X_2)$ and $Var(X_1)=Var(X_2)$, but we may have $X_1 \neq X_2$. Consequently, we can't write

$$X_1 + X_2 + ... + X_{100} = 100X \quad \text{(Wrong)}$$
$$Y_1 + Y_2 + ... + Y_{200} = 200Y \quad \text{(Wrong)}$$
$$S = 100X + 200Y \quad \text{(Wrong)}$$

If you mistakenly write $S = 100X + 200Y$, you'll get the wrong variance:

$$Var(S) = Var(100X + 200Y) = 100^2 Var(X) + 200^2 Var(Y) \quad \text{(Wrong)}$$

We know the correct formula is $Var(S) = 100Var(X) + 200Var(Y)$.

Let's continue. Now we need to find $E(X)$, $Var(X)$, $E(Y)$, and $Var(Y)$. Once again, we want to precisely specify X and Y:

Class	# of policies	Probability that each policy has a loss in a year	distribution of the annual loss per policy if there's a loss	
			Mean	Variance
1	100	0.1	1	3
2	200	0.3	2	5

$$\Rightarrow \quad X = \begin{cases} 0 & \text{with probability of } 0.9 \\ U & \text{with probability of } 0.1\text{, where } E(U)=1, \quad Var(U)=3 \end{cases}$$

$$\Rightarrow \quad Y = \begin{cases} 0 & \text{with probability of } 0.7 \\ V & \text{with probability of } 0.3\text{, where } E(V)=2, \quad Var(V)=5 \end{cases}$$

$$\Rightarrow \quad E(X) = 0.9(0) + 0.1E(U) = 0.1(1) = 0.1.$$

When calculating $Var(X)$, we have to be careful. We can't directly find $Var(X)$. For example, we can't write $Var(X) = 0.1 \times Var(U)$. We need to proceed using the formula $Var(X) = E(X^2) - E^2(X)$

$$E\left(X^2\right)=0.9\left(0^2\right)+0.1E\left(U^2\right)=0.1E\left(U^2\right)=0.1\left[E^2\left(U\right)+Var\left(U\right)\right]=0.1\left[1^2+3\right]=0.4$$

$$\Rightarrow \quad Var\left(X\right)=E\left(X^2\right)-E^2\left(X\right)=0.4-0.1^2=0.39$$

Similarly,

$$\Rightarrow \quad Y=\begin{cases}0 & \text{with probability of } 0.7\\ V & \text{with probability of } 0.3\text{, where } E\left(V\right)=2, \quad Var\left(V\right)=5\end{cases}$$

$$\Rightarrow \quad E\left(Y\right)=0.7\left(0\right)+0.3E\left(V\right)=0.3E\left(V\right)=0.3\left(2\right)=0.6$$

$$E\left(Y^2\right)=0.7\left(0^2\right)+0.3E\left(V^2\right)=0.3E\left(V^2\right)=0.3\left[E^2\left(V\right)+Var\left(V\right)\right]=0.3\left[2^2+5\right]=2.7$$

$$\Rightarrow \quad Var\left(Y\right)=E\left(Y^2\right)-E^2\left(Y\right)=2.7-0.6^2=2.34$$

$$\Rightarrow \quad E\left(S\right)=100E\left(X\right)+200E\left(Y\right)=100\left(0.1\right)+200\left(0.6\right)=130$$

$$\Rightarrow \quad Var\left(S\right)=100Var\left(X\right)+200Var\left(Y\right)=100\left(0.39\right)+200\left(2.34\right)=507$$

$$\Rightarrow \quad \frac{\sigma_S}{E\left(S\right)}=\frac{\sqrt{507}}{130}=0.1732$$

Normal approximation

Example 14

You are given information about a block of insurance policies:

Class	# of policies	Probability that each policy has a loss in a year	distribution of the annual loss per policy if there's a loss	
			Mean	Variance
1	100	0.1	3	4
2	400	0.2	5	6
3	500	0.3	7	8

Assume all these 1,000 policies are independent.

Let S represent the total annual loss incurred by the 1,000 policies.

- Calculate $\frac{\sigma_S}{E(S)}$, the coefficient of variation of S.
- Use normal approximation to calculate the probability that the total annual loss exceeds 95% of the expected total annual loss, that is $P\left[S > 95\% E(S)\right]$

Solution

Let X represent the annual loss incurred by a policy randomly chosen from Class 1. Let Y represent the annual loss incurred by a policy randomly chosen from Class 2. Let Z represent the annual loss incurred by a policy randomly chosen from Class 3.

$$S = X_1 + X_2 + ... + X_{100} + Y_1 + Y_2 + ... + Y_{400} + Z_1 + Z_2 + ... + Z_{500}.$$

$$\Rightarrow \quad E(S) = 100E(X) + 400E(Y) + 500E(Z)$$

$$\Rightarrow \quad Var(S) = 100Var(X) + 400Var(Y) + 500Var(Z)$$

Class	# of policies	Probability that each policy has a loss in a year	distribution of the annual loss per policy if there's a loss	
			Mean	Variance
1	100	0.1	3	4
2	400	0.2	5	6
3	500	0.3	7	8

$$\Rightarrow \quad X = \begin{cases} 0 & \text{with probability of } 0.9 \\ U & \text{with probability of } 0.1, \text{ where } E(U) = 3, \quad Var(U) = 4 \end{cases}$$

$$\Rightarrow \quad Y = \begin{cases} 0 & \text{with probability of } 0.8 \\ V & \text{with probability of } 0.2, \text{ where } E(V) = 5, \quad Var(V) = 6 \end{cases}$$

$$\Rightarrow \quad Z = \begin{cases} 0 & \text{with probability of } 0.7 \\ R & \text{with probability of } 0.3, \text{ where } E(R) = 7, \quad Var(R) = 8 \end{cases}$$

$$\Rightarrow \quad E(X) = 0.9(0) + 0.1E(U) = 0.1E(U) = 0.1(3) = 0.3$$

$$E\left(X^2\right) = 0.9\left(0^2\right) + 0.1E\left(U^2\right) = 0.1E\left(U^2\right) = 0.1\left[E^2(U) + Var(U)\right] = 0.1\left[3^2 + 4\right] = 1.3$$

$$\Rightarrow \quad Var(X) = E\left(X^2\right) - E^2(X) = 1.3 - 0.3^2 = 1.21$$

$$\Rightarrow \quad E(Y) = 0.8(0) + 0.2E(V) = 0.2E(V) = 0.2(5) = 1$$

$$E(Y^2) = 0.8(0^2) + 0.2E(V^2) = 0.2E(V^2) = 0.2\left[E^2(V) + Var(V)\right] = 0.2\left[5^2 + 6\right] = 6.2$$

$$\Rightarrow \quad Var(Y) = E(Y^2) - E^2(Y) = 6.2 - 1^2 = 5.2$$

$$\Rightarrow \quad E(Z) = 0.7(0) + 0.3E(R) = 0.3E(R) = 0.3(7) = 2.1$$

$$E(Z^2) = 0.7(0^2) + 0.3E(R^2) = 0.3E(R^2) = 0.3\left[E^2(R) + Var(R)\right] = 0.3\left[7^2 + 8\right] = 17.1$$

$$\Rightarrow \quad Var(Z) = E(Z^2) - E^2(Z) = 17.1 - 2.1^2 = 12.69$$

$$\Rightarrow \quad E(S) = 100E(X) + 400E(Y) + 500E(Z)$$
$$= 100(0.3) + 400(1) + 500(2.1) = 1,480$$

$$\Rightarrow \quad Var(S) = 100Var(X) + 400Var(Y) + 500Var(Z)$$
$$= 100(1.21) + 400(5.2) + 500(12.69) = 8,546$$

$$\Rightarrow \quad \frac{\sigma_S}{E(S)} = \frac{\sqrt{8,546}}{1,480} = 0.06246$$

Next, we need to find $P\left[S > 95\%E(S)\right]$. S is approximately normal.

$$P\left[S > 95\%E(S)\right] = 1 - P\left[S \le 95\%E(S)\right]$$

$$P\left[S \le 95\%E(S)\right] = \Phi\left[\frac{95\%E(S) - E(S)}{\sigma_S}\right] = \Phi\left[-5\%\frac{E(S)}{\sigma_S}\right] = \Phi\left[-5\%(16.01)\right]$$
$$= \Phi(-0.80) = 1 - \Phi(0.80)$$

$$P\left[S > 95\%E(S)\right] = 1 - P\left[S \le 95\%E(S)\right] = 1 - \left[1 - \Phi(0.8)\right] = \Phi(0.8) = 0.7881$$

Example 15

You are given information about a block of insurance policies:

Class	# of policies	Probability that each policy has a loss in a year	Loss amount if there's a loss
1	100	0.1	3
2	200	0.2	4

Assume all these 300 policies are independent.

Let S represent the total annual loss incurred by the 300 policies.

- Calculate $\dfrac{\sigma_S}{E(S)}$, the coefficient of variation of S.
- Use normal approximation to calculate the probability that the total annual loss exceeds 105% of the expected total annual loss, that is $P\left[S > 105\% E(S)\right]$

Solution

Let X represent the annual loss incurred by a policy randomly chosen from Class 1. Let Y represent the annual loss incurred by a policy randomly chosen from Class 2.

$S = X_1 + X_2 + ... + X_{100} + Y_1 + Y_2 + ... + Y_{200}$.

$\Rightarrow \quad E(S) = 100E(X) + 200E(Y)$

$\Rightarrow \quad Var(S) = 100Var(X) + 200Var(Y)$

$$X = \begin{cases} 0 & \text{with probability of } 0.9 \\ 3 & \text{with probability of } 0.1 \end{cases}$$

$$Y = \begin{cases} 0 & \text{with probability of } 0.8 \\ 4 & \text{with probability of } 0.2 \end{cases}$$

$E(X) = 0(0.9) + 3(0.1) = 0.3$

$E(X^2) = 0^2(0.9) + 3^2(0.1) = 0.9$

$Var(X) = E(X^2) - E^2(X) = 0.9 - 0.3^2 = 0.81$

$E(Y) = 0(0.8) + 4(0.2) = 0.8$

$E(Y^2) = 0^2(0.8) + 4^2(0.2) = 3.2$

$Var(Y) = E(Y^2) - E^2(Y) = 3.2 - 0.8^2 = 2.56$

To avoid the above calculation, you can use BA II Plus and quickly calculate the mean and variance for X and Y.

$\Rightarrow \quad E(S) = 100E(X) + 200E(Y) = 100(0.3) + 200(0.8) = 190$

$\Rightarrow \quad Var(S) = 100Var(X) + 200Var(Y) = 100(0.81) + 200(2.56) = 593$

$$\Rightarrow \quad \frac{\sigma_S}{E(S)} = \frac{\sqrt{593}}{190} = 0.1282$$

Next, we need to find $P\left[S > 105\% E(S)\right]$. S is approximately normal.

$$P\left[S > 105\% E(S)\right] = 1 - P\left[S \le 105\% E(S)\right]$$

$$P\left[S \le 105\% E(S)\right] = \Phi\left[\frac{105\% E(S) - E(S)}{\sigma_S}\right] = \Phi\left[5\% \frac{E(S)}{\sigma_S}\right] = \Phi\left[5\%(7.8)\right]$$
$$= \Phi(0.39) = 0.6517$$

$$P\left[S > 105\% E(S)\right] = 1 - P\left[S \le 105\% E(S)\right] = 1 - 0.6517 = 0.3483$$

Example 16

Two random loss variables, X and Y have the following joint density function:

$$f_{X,Y}(x,y) = \frac{4}{81} xy, \qquad \text{where } 0 \le X \le 3 \text{ and } 0 \le Y \le 3$$

The insurer pays $X+Y$ in excess of deductible 1, subject to the maximum payment of 2.

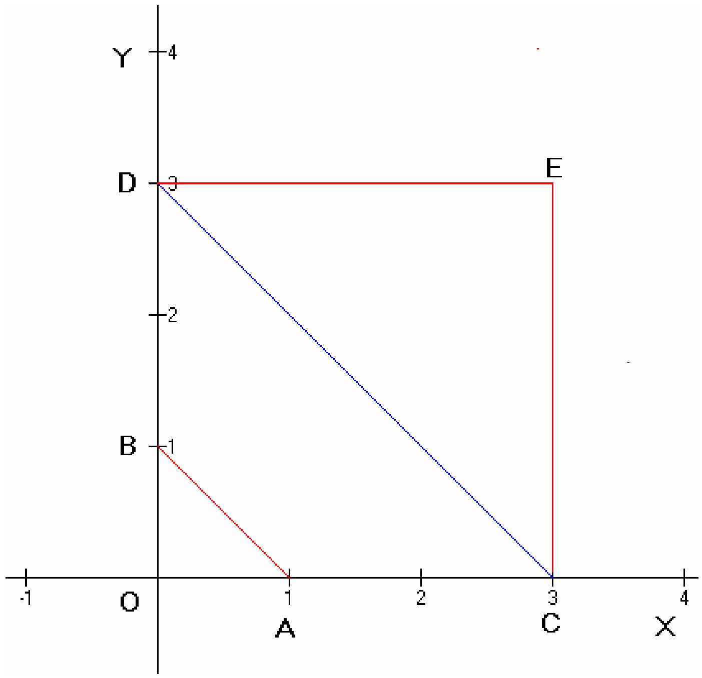
Calculate the expected claim payment by the insurer.

Solution

Let S represent the claim payment by the insurer. Then,

$$S = \begin{cases} 0 & \text{if } \ 0 \le X+Y \le 1 \\ X+Y-1 & \text{if } \ 1 \le X+Y \le 3 \\ 2 & \text{if } \ 3 \le X+Y \end{cases}$$

Next, we draw a diagram



In the above diagram, the square ODEC is where X and Y exist. AB represents $x+y=1$. CD represents $x+y=3$.

Area AOB is where $0 \le x+y \le 1$. When (x,y) falls in AOB, the insurer pays nothing. Area ABDC is where $1 \le x+y \le 3$; here the insurer pays $x+y-1$. Area CDE is where $3 \le x+y$, $x \le 3$, and $y \le 3$; here the insurer pays 2.

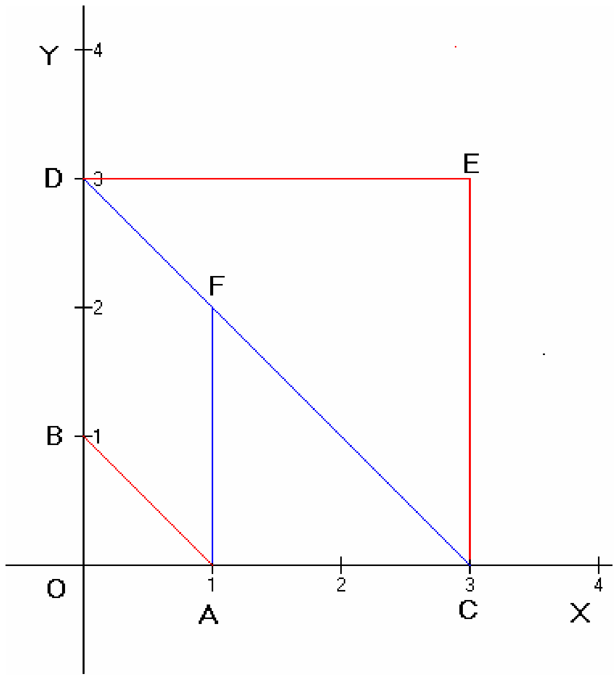
$$S = \begin{cases} 0 & \text{if} \quad 0 \le X+Y \le 1 \\ X+Y-1 & \text{if} \quad 1 \le X+Y \le 3 \\ 2 & \text{if} \quad 3 \le X+Y \end{cases}$$

$$E(S) = \iint_{AOB} 0 f(x,y)\,dxdy + \iint_{ABDC} (x+y-1) f(x,y)\,dxdy + \iint_{CDE} 2 f(x,y)\,dxdy$$

$$\iint_{AOB} 0 f(x,y)\,dxdy = 0$$

$$\iint_{ABDC} (x+y-1) f(x,y)\,dxdy = \iint_{ABDC} (x+y-1)\frac{4}{81} xy\,dx\,dy.$$

To do this integration, we divide ABDC into two areas: ABDF and AFC.



$$\iint_{ABDC} (x+y-1)\frac{4}{81}xy\,dx\,dy = \iint_{ABDF} (x+y-1)\frac{4}{81}xy\,dx\,dy + \iint_{AFC} (x+y-1)\frac{4}{81}xy\,dx\,dy$$

$$\iint_{ABDF} (x+y-1)\frac{4}{81}xy\,dx\,dy = \int_0^1\left[\int_{1-x}^{3-x}(x+y-1)\frac{4}{81}xy\,dy\right]dx = \int_0^1\left(\frac{56}{243}x - \frac{8}{81}x^2\right)dx = \frac{20}{243}$$

$$\iint_{AFC} (x+y-1)\frac{4}{81}xy\,dx\,dy = \int_1^3\left[\int_0^{3-x}(x+y-1)\frac{4}{81}xy\,dy\right]dx = \frac{184}{1215}$$

$$\iint_{ABDC} (x+y-1)\,f(x,y)\,dxdy = \frac{20}{243} + \frac{184}{1215} = 0.2337$$

$$\iint_{CDE} 2f(x,y)\,dxdy = \int_0^3\left[\int_{3-x}^{3} 2\left(\frac{4}{81}xy\right)dy\right]dx = \frac{5}{3} = 1.6667$$

$$\Rightarrow \quad E(S) = \iint_{ABDC} (x+y-1)\,f(x,y)\,dxdy + \iint_{CDE} 2f(x,y)\,dxdy = 0.2337 + 1.6667 = 1.90$$

Example 17

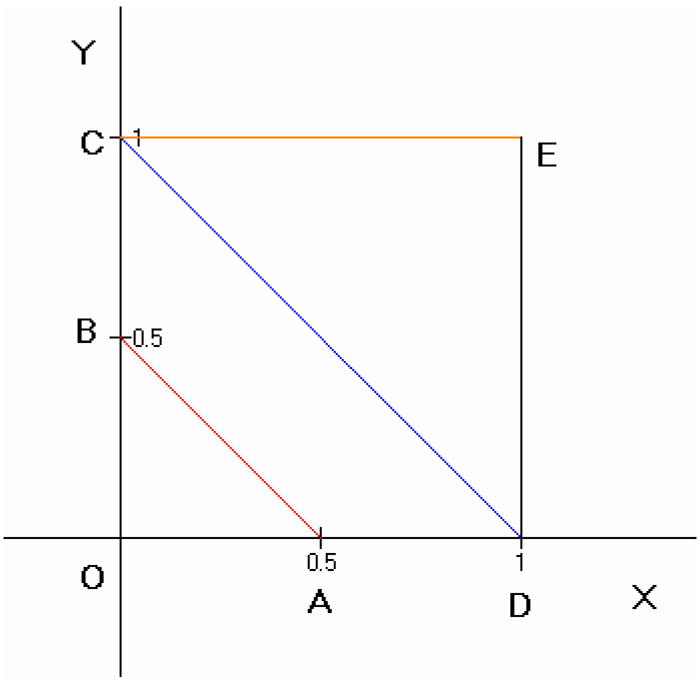
Two random loss variables, X and Y have the following joint density function:

$$f_{X,Y}(x,y) = 2y, \quad \text{where } 0 \le X \le 1 \text{ and } 0 \le Y \le 1$$

The insurer pays $X + Y$ in excess of deductible 0.5, subject to the maximum payment of 0.5.

Calculate the expected claim payment by the insurer.

Solution



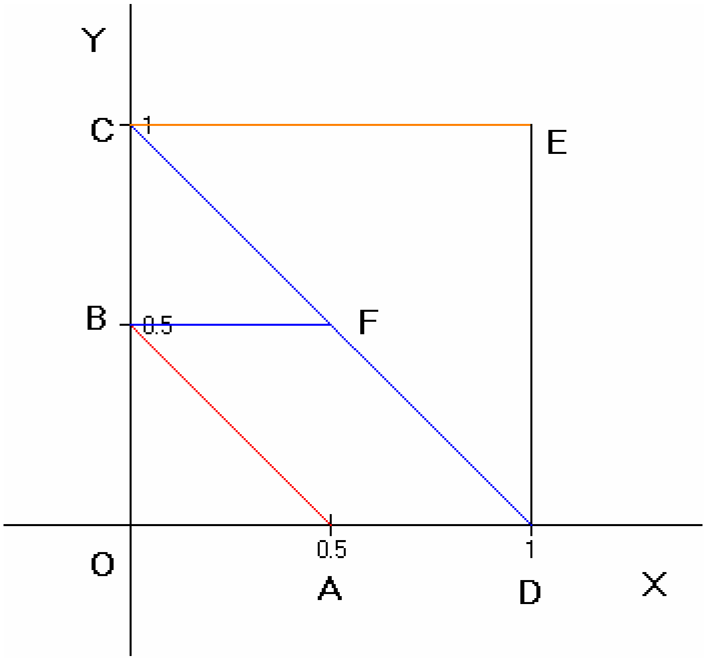
Let S represent the claim payment by the insurer. Then,

$$S = \begin{cases} 0 & \text{if } \ 0 \le X + Y \le 0.5 \\ X + Y - 0.5 & \text{if } \ 0.5 \le X + Y \le 1 \\ 0.5 & \text{if } \ 1 \le X + Y \end{cases}$$

$$E(S) = \iint_{AOB} 0 f(x,y)\,dxdy + \iint_{ABCD} (x + y - 0.5) f(x,y)\,dxdy + \iint_{CDE} 0.5 f(x,y)\,dxdy$$

$$= \iint_{ABCD} (x+y-0.5) f(x,y)dxdy + 0.5 \iint_{CDE} f(x,y)dxdy$$

To calculate $\iint_{ABCD} (x+y-0.5) f(x,y)dxdy$, we divide ABCE into two sub-areas: ABFD and BCF.



$$\iint_{ABCD} (x+y-0.5) f(x,y)dxdy$$

$$= \iint_{ABFD} (x+y-0.5) f(x,y)dxdy + \iint_{BCF} (x+y-0.5) f(x,y)dxdy$$

$$\iint_{ABFD} (x+y-0.5) f(x,y)dxdy = \int_0^{0.5} \left[\int_{0.5-y}^{1-y} (x+y-0.5) f(x,y)dx \right] dy$$

$$= \int_0^{0.5} \left[\int_{0.5-y}^{1-y} (x+y-0.5) 2y\, dx \right] dy = \int_0^{0.5} \frac{1}{8} ydy = \frac{1}{64}$$

$$\iint_{BCF} (x+y-0.5) f(x,y)dxdy = \int_0^{0.5} \left[\int_{0.5}^{1-x} (x+y-0.5) 2ydy \right] dx = 0.05729$$

$$\Rightarrow \iint_{ABCD} (x+y-0.5) f(x,y)dxdy = \frac{1}{64} + 0.05729 = 0.072915$$

$$0.5\iint_{CDE} f(x,y)\,dxdy = 0.5\int_0^1\int_{1-x}^1 2y\,dydx = 0.5\times\frac{2}{3} = \frac{1}{3}$$

$$\Rightarrow \quad E(S) = 0.072915 + \frac{1}{3} = 0.40625$$

Security loading

Example 18

Losses are uniformly distributed over [0,4000]. The insurer will pay the loss amount in excess of the deductible of 1,000.

The insurer, in setting the premium rate, has a security loading of 10%.

Calculate the gross premium.

Solution

This study note talks about security loading θ. When an insurer sets its premium rate, it can't set it just equal to the expected loss. The insurer needs to charge more to pay expense and earn a profit.

Let P represent the gross premium. Let Y represent the claim payment. Then,

$$P = (1+\theta)E(Y)$$

Let X represent the loss and Y the payment made by the insurer. Then $Y = 0$ if $X \le 1000$; $Y = X - 1000$ if $X \ge 1000$. We need to first calculate $E(Y)$.

$$E(Y) = \int_0^{1000} 0 f(x)dx + \int_{1000}^{4000}(x-1000)f(x)dx = \int_{1000}^{4000}(x-1000)\frac{1}{4000}dx$$

Set $x - 1000 = t$.

$$\int_{1000}^{4000}(x-1000)\frac{1}{4000}dx = \int_0^{3000} t\frac{1}{4000}dt = \frac{1}{4000}\times\frac{1}{2}\left[t^2\right]_0^{3000} = 1{,}125$$

$$\Rightarrow \quad P = (1+\theta)E(Y) = (1+10\%)1125 = 1{,}237.5$$

Chapter 37 On becoming an actuary…

As you contemplate entering the actuarial field, several questions resurface across the board from candidates. Let's explore together what challenges, obstacles, and opportunities you might encounter along the way.

Why do people pursue the actuarial career track?

- Bypass an expensive MBA, JD, or MD degree.
- Earn what you learn; you get a raise every time you pass an exam. Diligent candidates can reap rich monetary rewards for their hard work by their late 20's or early 30's.
- A steady market demand for actuaries and the highly technical nature of the profession provide a job security few careers can equal in today's volatile workplace.
- White collar work environment.

What is the biggest challenge candidates face as they work toward becoming a fellow?

EXAMS! Several hundred study hours are typically needed to pass one exam. Becoming a fellow requires you to study for and pass a total of eight exams. You must expect to lead the life of a professional student for several years until you have passed all the exams.

A key character quality that candidates must possess in abundance is perseverance. Sometimes it may take you one or more attempts to finally pass one exam. If you fail, you need to find the motivation to study again for the same exam. Studying for the same exam and reading the same textbooks over and over can be discouraging.

Lastly, exams add a lot of pressure to an already full professional and personal life, especially if you are married and have children. If you are single, try to pass as many exams as you can while you are single. If you are married and have children, you'll need your spouse's support to free up more study time, particularly when the exam is only a few weeks away.

How many exams do I need to pass to get a job?

If you are a young college student, Exam P alone may open the door for employment. Older career switchers or international candidates may have to pass more exams to win a potential employer's attention.

Each candidate must weigh his own unique skill set and determine the delicate balance of book knowledge versus job experience. Having more exams under your belt may impress an employer and hence open more doors. Passing exams is also an excellent confidence-builder. However, don't pass too many exams (such as passing 5 or 6 exams). Employers want well-balanced candidates whose exam levels are in line with their work experience.

I'm an older candidate who wants to make a switch to the actuary field. How can I take the leap?

Many companies target college students for their entry-level positions so that by the time they are in their late 20's they become a manager and then perhaps move to director by their mid 30's. Older candidates, of course, will not follow this traditional path. In talking with potential employers, older candidates will want to focus on the unique skills and talents as well as the added maturity and responsibility they can bring to bear on their jobs. Study diligently and excel at demonstrating your skills of not only passing exams but also bringing your knowledge to bear on your daily job experience.

Another possibility for making the leap to a new career in the actuary field is to first get a job as a secretary, technical assistant, or programmer in an actuary department. Once you are hired, then you can study for exams and try to become an actuary.

What computer skills do employers look for in their actuarial job candidates?

The most commonly used software programs in the actuarial field are Excel and Access. Learning how to program in one or both of these will give your resume an extra boost.

Should I specialize in P&C (property and casualty) or life?

As a newcomer to the actuary field, remain open to both areas until you have worked as an actuary for several years. As you gain experience, your skills and interests will probably point to the area of specialization which fits you best.

Guo's Mock Exam

Allotted time: 180 minutes

To best use this mock exam, please print this PDF, find a quiet place, take this exam under the strict exam-like condition.

Turn to the next page.

Q1
Random variable X has a Poisson distribution with mean $\lambda = 3.2$. Calculate the mode of X.

A 2 B 3 C 4 D 5 E 6

Q2
Random variable X has the following pdf:

$$f(x) = 3x^2, \text{ where } 0 < x < 1$$

You take a random sample of 3 with median Y. Find the pdf of Y.

A $18y^5(1-y^3)$ B $4y(1-y^2)$ C $12y^2(1-y)$
D $6y^2(1-y^3)$ E $30y^4(1-y)$

Q3
An insurance company divides its large pool of policyholders into three groups: standard class, preferred class, and super preferred class. You are given the following information:

- There are 16 times as many policyholders in the standard risk as in the super preferred class
- The # of policyholders in the super preferred class is one third of the policyholders in the preferred class
- The probability that a super preferred policyholder dies next year is one-sixth of the probability that a standard policyholder dies next year
- The probability that a preferred policyholder dies next year is two thirds of the probability that a standard policyholder dies next year

Calculate the conditional probability that a policyholder is a super preferred given he dies next year.

A 0.5% B 1% C 1.5% D 2% E 2.5%

Q4
A multi-line insurer sells 3 types of insurance polices: auto, home, and life. Policyholders who have two or three types of policies get a multi-line discount. Due to discount, the policyholders who have all three policies have 85% chance of renewing their policies next year. In contrast, those who have only two policies and only one policy have 70% and 50% chance of renewing their policies next year.

You are given:

- All policyholders have at least one policy
- 51% of the policyholders have home policies
- 37% of the policyholders have life policies
- 15% of the policyholders have all three policies
- 22% of the policyholders have both auto and home polices
- 17% of the policyholders have both auto and life policies
- 19% of the policyholders have both home and life policies

Calculate the conditional probability that a randomly selected policyholder has only auto insurance given he renews his policy or policies next year.

A 0.13 B 0.15 C 0.17 D 0.22 E 0.27

Q5
An insurance company sells a special decreasing life insurance policy. The policy provides $3,000 death benefit if the insured dies in Year 1. The death benefit decreases each year by $5,000. The conditional probability that the insured dies each year given he's still alive at the beginning of the year is 0.2.

Calculate the pure premium of this policy.
A 14,875 B 15,000 C 15,240 D 15,750 E 16,200

Q6

X is the loss random variable with density function $f(x)=\frac{1}{4}e^{-\frac{x}{4}}$.
Z is the portion of the loss not covered by the insurance. Z is equal to 1 with probability of 0.4 and equal to 0 with probability of 0.6.

X and Z are independent.

Calculate $Var(XZ)$.

A 7.24 B 8.24 C 9.24 D 10.24 E 11.24

Q7

Q7 MGF for a loss random variable X is $M_x(t)=\frac{0.2}{0.2-t}$. The payment is $Y=80\%X+10$. What's the MGF for the payment?

A $\frac{1}{1-4t}e^{10t}$ B $\frac{0.2}{0.2-t}e^{10t}$ C $\frac{0.2}{0.2-t}e^{0.8t}$ D $\frac{1}{1-4t}e^{0.8t}$ E $\frac{1}{1-0.2t}e^{10t}$

Q8
The joint pdf is f(x,y)=2|x|y where -1<x<1 and 0<y<1. Calculate $E\left(X^2Y\right)$.

A 0.0 B 0.25 C 0.33 D 0.50 E 0.75

Q9
April 15 is approaching. Two taxpayers, Adam and Bob, plan to visit the same IRS office in town to ask for tax questions. Adam's arrival time to the IRS office is uniformly distributed over [10:00 am, 12:30 pm]. Bob's arrival time to the IRS office is uniformly distributed over [11:00 am, 1:00 pm]. Adam's arrival time to the IRS office is independent from Bob's arrival time to the IRS office.

The IRS office hours are from 8:00am to 4:00 pm except for a 30-minute lunch break, which lasts from 11:30 am to noon.

X = The probability that both Adam and Bob have to wait for no more than 20 minutes before they can approach an IRS clerk to ask for a question.

Y = The probability that both Adam and Bob have to wait for no more than 20 minutes before they can approach an IRS clerk to ask for a question given they are already waiting.

Calculate X + Y.

A 0.3 B 0.5 C 0.8 D 1.0 E 1.3

Q10
$$f\left(x,y\right)=\frac{5\left(y-x\right)^4\left(y-2\right)\left(4-y\right)}{4y^4} \quad \text{where } 0<x<y \text{ and } 2<y<4.$$
Calculate $E\left(X\right)$

A 0.00 B 0.11 C 0.28 D 0.51 E 0.75

Q11, Q12, Q13, Q14
The liability that results from a car accident falls into two categories: the property damage liability and the personal injury liability. Let random variables X and Y represent the dollar amount (in $10,000) of the property damage liability and the personal injury liability respectively. The pdf of X is:

$$f_X(x) = \frac{1}{8}x, \text{ where } 0 \le x \le k \text{ and } k \text{ is a constant}$$

Given $X = x$, Y is uniformly distributed over $[x, 2x]$.

An insurance policy is written to cover the sum of X and Y. There's a deductible of $30,000 and security loading of 30%.

Q11 The probability of having an accident where the property damage liability exceeds $20,000 and the personal injury liability exceeds $30,000.
A 0.688 B 0.712 C 0.733 D 0.752 E 0.801

Q12 Calculate the probability that insurer doesn't incur any claims.
A 0.078 B 0.082 C 0.094 D 0.102 E 0.124

Q13 Calculate the gross premium
A 4,000 B 4,300 C 4,600 D 4,900 E 5,200

Q14 The expected non-zero claim payment
A 3,250 B 3,750 C 3,985 D 4,000 E 4,150

Q15 You take a bus to work every workday. Your journey to work consists of 3 independent components:

- The time for you to walk from home to the bus stop nearby is normally distributed with mean of 5 minutes and standard deviation of 2 minutes.

- Your bus arrives immediately after you get to the bus stop. The time for a bus to take you to the stop near your office is normally distributed with mean of 20 minutes and 6 minutes.

- The time for you to walk to work from the bus stop to your company is normally distributed with mean of 3 minutes and standard deviation of 1 minute.

Your boss drives to work every workday. The time for him to drive to work is normally distributed with mean of 26 minutes with standard deviation of 5 minutes.

Calculate the probability that in a workday you arrive at work at least 5 minutes earlier than you boss does (assuming you leave your home and your boss leaves his home at the same time).

A 13% B 19% C 25% D 31% E 37%

Q16
A manufacturing plant purchases a special product defect insurance policy. The insurance provides a payment of $1,000 if the number of defects is 20. Then for each full 5 incremental defects, the insurance pays an additional $500. However, the payment by the insurance policy can be no more than $2,800 regardless of the number of defects.

The probability is 0.19 that the plant will have less than 5 defects.
The probability is 0.36 that the plant will have less than 10 defects.
The probability is 0.51 that the plant will have less than 15 defects.
The probability is 0.64 that the plant will have less than 20 defects.
The probability is 0.75 that the plant will have less than 25 defects.
The probability is 0.84 that the plant will have less than 30 defects.
The probability is 0.91 that the plant will have less than 35 defects.
The probability is 0.96 that the plant will have less than 40 defects.
The probability is 0.99 that the plant will have less than 45 defects.
The probability is 1.00 that the plant will have less than 50 defects.

Let Y represent the payment by the insurance policy. Calculate $\frac{E(Y)}{\sigma(Y)}$, the ratio of the mean to the standard deviation of Y.

A 0.48 B 0.53 C 0.58 D 0.63 E 0.68

Q17

X is a Poisson random variable. $\frac{F(2)}{F(1)} = 2.125$. Find $E(X)$

A 3 B 4 C 5 D 6 E 7

Q18

Random variables X and Y are jointly uniformly distributed over the area $0 \le |X| + |Y| \le 1$. Calculate σ_{XY}, the standard deviation of XY.

A 0.000 B 0.016 C 0.032 D 0.105 E 0.250

Q19

Random variables X,Y,Z have the following joint pdf: f(x,y,z)=$\frac{1}{\sqrt{xy}}$ where 0<x<y<z<1.

Calculate the probability that at least two of the three random variables X,Y,Z are less than 0.1.

A 0.20 B 0.25 C 0.30 D 0.35 E 0.40

Q20 In a small town, 55% are men and 45% are women. 74% of the women read the local newspaper everyday. Given that someone reads the local newspaper everyday, the probability that the reader is a male is 57.53%. Calculate the percentage of the men who read the local newspaper everyday.

A 0.78 B 0.82 C 0.88 D 0.92 E 0.98

Q21

A wife and husband's lives are uniformly distributed on (0,20) in years. Find the conditional probability that the husband outlives the wife given that the husband is still alive 10 years from today.

A 0.55 B 0.65 C 0.75 D 0.85 E 0.95

Q22
X is exponentially distributed with mean of 1. $Y=|X-4|$. Calculate $f_Y(2)$, the pdf of Y at $y=2$.

A 0.14 B 0.17 C 0.21 D 0.25 E 0.31

Q23
Random variable X has the following moment generating function:

$$M_X(t)=(1-2t)^{-\frac{1}{2}}$$

Calculate the coefficient of variation.

A 0.2 B 0.5 C 0.8 D 1.1 E 1.4

Q24
Random loss variables X and Y have the following joint pdf:

$$f(x,y)=x+k\,y^2\text{, where } 0\le x\le y\le 1$$

An insurance policy is written to cover $X+Y$. The maximum claim payment is \$1. Calculate the net premium

A 0.\$75 B 0.\$94 C \$1.0 D \$1.6 E \$2.2

Q25
A device has 4 duplicate components working in parallel. The device works as long as at least one component works; it fails only if all four components fail simultaneously. The times-to-failure of the 4 components are independent exponential random variables with mean of 1, 2, 3, and 4 hours respectively.

Calculate the probability that the device is still working 5 hours later.

A 0.32 B 0.37 C 0.42 D 0.47 E 0.52

Q26
Loss random variable X has the following pdf:

$$f(x) = 0.02x \text{, where } 0 < x < 10$$

The insurer has a deductible of 4 per loss.

Calculate the expected claim payment net of deductible.

A 2.5 B 2.7 C 2.9 D 3.4 E 4.2

Q27
Two discrete random variables X and Y are jointly distributed over a series of points. The joint probability mass function is $p_{X,Y}(x, y) = \frac{1}{28}$. You are also given:

- X =0, 1, 2, 3, 4, 5, 6
- Y =0, …, X (i.e. for each value in X, Y is a non-negative integer ranging from 0 to X)

Calculate $\frac{Var(X)}{Var(Y)}$

A 0 B 0.5 C 1 D 1.5 E 2

Q28
A company pays a benefit of 100 for each of its 1000 employees if an employee dies next year. The probability that an employee dies next year is 2%. What is the amount the company needs to have in a fund in order to ensure a 95% chance that it can cover the loss?

A 2,500 B 2,800 C 3,100 D 3,400 E 3,700

Q29
Random variables X and Y have the following joint pdf:

$f(x, y) = k\dfrac{x}{y}$, where $0 < x < y < 1$ and k is a constant.

Calculate $E\left(X \middle| Y = \dfrac{2}{3}\right) + Var\left(X \middle| Y = \dfrac{2}{3}\right)$

A 0.17 B 0.27 C 0.37 D 0.47 E 0.57

Q30
Random variable X has the following moment generating fuction: $M_X(t) = \dfrac{1}{1-\theta t}$, where θ is a positive constant. Calculate the probability that X is one standard deviation from its mean.

A 0.46 B 0.56 C 0.66 D 0.76 E 0.86

Solution to Guo's Mock Exam

Problems	Keys
1	B
2	A
3	B
4	E
5	C
6	D
7	A
8	C
9	E
10	D
11	A
12	C
13	D
14	E
15	B
16	E
17	A
18	D
19	A
20	B
21	C
22	A
23	E
24	B
25	D
26	C
27	C
28	B
29	D
30	E
A	6
B	6
C	6
D	6
E	6

If you answered at least 20 questions correctly, you passed this exam.

Q1
Random variable X has a Poisson distribution with mean $\lambda = 3.2$. Calculate the mode of X.

A 2 B 3 C 4 D 5 E 6

Solution **B**

$p_X(x) = e^{-\lambda}\dfrac{\lambda^x}{x!}$ where $x = 0, 1, 2, ..., n, ...$

MOde = Most Often (Most Observed). At the mode, $p_X(x)$ reaches its max value.

If X is continuous, we'll find the mode by solving the equation $\dfrac{d}{dx} f_X(x) = 0$. If X is discrete, we can't set $\dfrac{d}{dx} p_X(x) = 0$. The general strategy for finding the mode for a discrete random variable is to look at the ratio $\dfrac{p(x)}{p(x-1)}$.

In this case, $p(x-1) = e^{-\lambda}\dfrac{\lambda^{x-1}}{(x-1)!}$, $p_X(x) = e^{-\lambda}\dfrac{\lambda^x}{x!}$ $\Rightarrow \dfrac{p(x)}{p(x-1)} = \dfrac{e^{-\lambda}\dfrac{\lambda^x}{x!}}{e^{-\lambda}\dfrac{\lambda^{x-1}}{(x-1)!}} = \dfrac{\lambda}{x}$

So $p(x) \geq p(x-1)$ if $\dfrac{\lambda}{x} \geq 1$ or $x \leq \lambda$. For $x = 0, 1, 2, ...,$ $p(x)$ will keep increasing as x increases until $x = \text{int}(\lambda)$. Here $\text{int}(\lambda)$ means the integer portion of λ :

$$\text{int}(\lambda) = \text{int}(3.2) = 3$$

At $x = \text{int}(\lambda)$, $p(x)$ reaches its peak. Then $p(x)$ declines as x exceeds $\text{int}(\lambda)$. Here $\text{int}(\lambda)$ represents the integer portion of λ. So here the mode is $\text{int}(3.2) = 3$. Let's test a few values:

$X = x$	$p_X(x) = e^{-3.2}\dfrac{3.2^x}{x!}$
0	0.040762
1	0.130439
2	0.208702
3	**0.222616**
4	0.178093
5	0.113979

As a bonus point, let's find the mode of a binomial random variable x with parameter n and p.

$P(x)=C_n^x p^x q^{n-x}$, $P(x-1)=C_n^{x-1} p^{x-1} q^{n-x+1}$

$$\Rightarrow \qquad \frac{P(x)}{P(x-1)}=\frac{C_n^x p^x q^{n-x}}{C_n^{x-1} p^{x-1} q^{n-x+1}}=\frac{n-x+1}{x}\times\frac{p}{q}$$

Set $\frac{n-x+1}{x}\times\frac{p}{q}\geq 1$.

$\Rightarrow$ $p(n-x+1)\geq qx$, $pn-px+p\geq qx$, $px+qx\leq p(n+1)$

However, $p+q=1$. So we have: $x\leq p(n+1)$.

The mode of a Poisson random variable is $\text{int}(\lambda)$. The mode of a binominal random variable X is $\text{int}\left[p(n+1)\right]$.

Q2
Random variable X has the following pdf:

$$f(x)=3x^2\text{, where } 0<x<1$$

You take a random sample of 3 with median Y. Find the pdf of Y.

A $18y^5(1-y^3)$ B $4y(1-y^2)$ C $12y^2(1-y)$
D $6y^2(1-y^3)$ E $30y^4(1-y)$

Solution A

$$f(x)=3x^2,\quad P(X\leq x)=\int_0^x f(t)dt=\int_0^x 3t^2dt=x^3,\quad P(X>x)=1-x^3$$

Of the 3 samples taken, one of the 3 samples must fall in the range of $(y,y+dy)$, one of the 2 remaining samples must be greater than $y+dy$, and the 3rd sample must be smaller than y. This way, we are guaranteed that the median of the 3 samples is falls in the range $(y, y+dy)$.

$$f_Y(y)dy = \underbrace{3!}_{\text{Permutation}} \underbrace{P(X > y+dy)}_{\text{one sample must be greater than } y} \underbrace{P(y < X < y+dy)}_{\text{one sample must fall in } (y, y+dy)} \underbrace{P(X < y)}_{\text{one sample must be less than } y}$$

$P(X > y+dy) = P(X > y)$ because dy is tiny

$= 1 - y^3$

$P(X < y) = y^3$, $\quad P(y < X < y+dy) = f_X(y)dy = 3y^2 dy$

$\Rightarrow \quad f_Y(y)dy = 3!(1-y^3)(3y^2dy)(y^3) = 18y^5(1-y^3)dy$, where $0 < y < 1$

$\Rightarrow \quad f_Y(y) = 18y^5(1-y^3)$, where $0 < y < 1$

Double check:

$$\int_0^1 f_Y(y)dy = \int_0^1 18y^5(1-y^3)dy = 18\int_0^1 (y^5 - y^8)dy = 18\left[\frac{1}{6}y^6 - \frac{1}{9}y^3\right]_0^1 = 18\left[\frac{1}{6} - \frac{1}{9}\right] = 3 - 2 = 1$$

So $f_Y(y) = 18y^5(1-y^3)$ is the legitimate pdf having a total probability of one.

Q3

An insurance company divides its large pool of policyholders into three groups: standard class, preferred class, and super preferred class. You are given the following information:

- There are 16 times as many policyholders in the standard risk as in the super preferred class
- The # of policyholders in the super preferred class is one third of the policyholders in the preferred class
- The probability that a super preferred policyholder dies next year is one-sixth of the probability that a standard policyholder dies next year
- The probability that a preferred policyholder dies next year is two thirds of the probability that a standard policyholder dies next year

Calculate the conditional probability that a policyholder is a super preferred given he dies next year.

A 0.5% B 1% C 1.5% D 2% E 2.5%

Solution B

Let x represent the portion of the policyholders that are in the super preferred class. Let y represent the probability that a super preferred policyholder dies next year. To solve the problem, we don't need to know x or y (though we can calculate x). Please note we don't have enough info to calculate y.

Event: A policyholder dies next year

Segment	Segment's size	Segment's prob	Segment's contribution	Segment's contribution %
Standard	$16x$	$6y$	$16x(6y)$ $=96xy$	$\frac{96xy}{109xy}=88.07\%$
Preferred	$3x$	$6y\left(\frac{2}{3}\right)=4y$	$3x(4y)$ $=12xy$	$\frac{12xy}{109xy}=11.01\%$
Super Preferred	x	y	$x(y)$ $=xy$	$\frac{xy}{109xy}=0.92\%$
Total	100%		$(96+12+1)xy$ $=109xy$	100%

The conditional probability that a policyholder is a super preferred given he dies next year is 0.92%.

Q4

A multi-line insurer sells 3 types of insurance polices: auto, home, and life. Policyholders who have two or three types of policies get a multi-line discount. Due to discount, the policyholders who have all three policies have 85% chance of renewing their policies next year. In contrast, those who have only two policies and only one policy have 70% and 50% chance of renewing their policies next year.

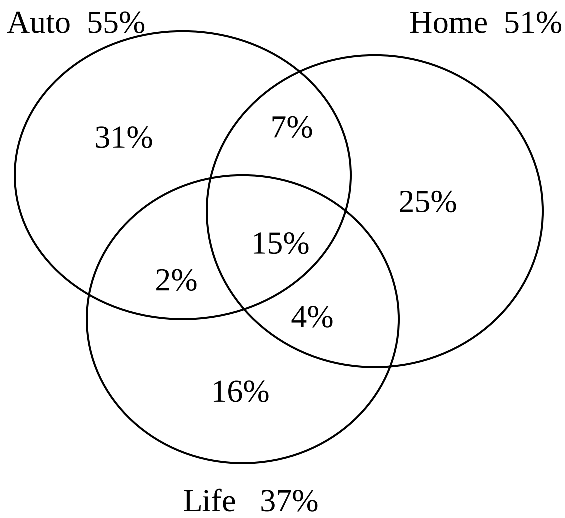
You are given:

- All policyholders have at least one policy
- 51% of the policyholders have home policies
- 37% of the policyholders have life policies
- 15% of the policyholders have all three policies
- 22% of the policyholders have both auto and home polices
- 17% of the policyholders have both auto and life policies
- 19% of the policyholders have both home and life policies

Calculate the conditional probability that a randomly selected policyholder has only auto insurance given he renews his policy or policies next year.

A 0.13 B 0.15 C 0.17 D 0.22 E 0.27

Solution E



Event: A random selected policyholder renews his policy (policies) next year

Segment	Segment size	Segment's probability to make the event happen	Segment's contribution	Segment's contribution %
Auto only	31%	50%	31%(50%)=0.1550	0.1550/0.5785=26.79%
Home only	25%	50%	25%(50%)=0.1250	0.125/0.5785=21.61%
Life only	16%	50%	16%(50%)=0.0800	0.08/0.5785=13.83%
Two policies only	13%	70%	13%(70%)=0.0910	0.091/0.5785=15.73%
Three policies only	15%	85%	15%(85%)=0.1275	0.1275/0.5785=22.04%
Total	100%		0.5785	100.00%

The probability is 27% that a randomly selected policyholder has only auto insurance given he renews his policy or policies next year.

Q5
An insurance company sells a special decreasing life insurance policy. The policy provides $30,000 death benefit if the insured dies in Year 1. The death benefit decreases each year by $5,000. The conditional probability that the insured dies each year given he's still alive at the beginning of the year is 0.2.

Calculate the pure premium of this policy.

A 14,875 B 15,000 C 15,240 D 15,750 E 16,200

Solution C

The pure premium is the expected death benefit.

Year	Death benefit	Probability of death that year
1	30,000	0.2
2	25,000	0.8*0.2
3	20,000	$0.8^2 * 0.2$
4	15,000	$0.8^3 * 0.2$
5	10,000	$0.8^3 * 0.2$
6	5,000	$0.8^4 * 0.2$

For example, this is how to find the probability that the insured dies in Year 2. For him to die in Year 2, he must (1) be alive at Year 1 (with probability 0.8), (2) die in Year 2 given he's alive in Year 1 (probability 0.2). So the probability that the insured dies in Year 2 is 0.8*0.2. Mathematically, this is:

$$P(\text{die in Year 2})$$
$$= P(\text{alive @ end of Year 1})P(\text{die in Year 2}|\text{alive @end of Year 1}) = 0.8 \times 0.2$$

The expected death benefit (in $1,000):

$$30(0.2)+25(0.8*0.2)+20(0.8^2 * 0.2)+15(0.8^3 * 0.2)+10(0.8^3 * 0.2)+5(0.8^4 * 0.2)$$
$$=15.24288$$

So the pure premium is $15,243.

Since this problem doesn't ask you to calculate variance of the payment, you don't need to use BA II Plus 1-V Statistics Worksheet. If the problem asks you to also find the standard deviation, then you might want to use BA II Plus 1-V Statistics Worksheet. However, you'll need to be very careful if you do use BA II Plus 1-V Statistics Worksheet: you'll need to include the probability of zero payment at the end of Year 6.

The total probability that the policyholder dies in a 6-year period is
0.2+0.16+0.128+0.1024+0.08192+0.065536 = 0.737856.

The probability that he is still alive at the end of Year 6 is:
1 - 0.737856 = 0.262144.

A quick way to calculate the probability of being alive at the end of Year 6:
0.8^6=0.262144

Next, you can set up the following table:

Year	Death benefit	Probability of getting this death benefit	Scale up probability: multiply the probability by 10,000
1	30,000	0.2	200
2	25,000	0.8*0.2=0.16	160
3	20,000	$0.8^2 * 0.2$=0.128	128
4	15,000	$0.8^3 * 0.2$=0.1024	102
5	10,000	$0.8^3 * 0.2$=0.08192	82
6	5,000	$0.8^4 * 0.2$=0.065536	66
6	0	0.8^6=0.262144	262
Total		1.0	10,000

Next, enter the following in 1-V Statistics Worksheet:

Year	Death benefit (setting $1,000 as one unit of money to speed up the calculation)	Scale up probability: multiply the probability by 1,000
1	X01=30	Y01=200
2	X02=25	Y02=160
3	X03=20	Y03=128
4	X04=15	Y04=102
5	X05=10	Y05=82
6	X06=5	Y06=66
6	X07=0	Y07=262
Total		1,000

You should get: $E(X)$=15.24=$15,240, σ_X =11.47790922=$11,477.91

If you forget to include the zero payment, your table will becomes:

Year	Death benefit (setting $1,000 as one unit of money to speed up the calculation)	Scale up probability: multiply the probability by 1,000
1	X01=30	Y01=200
2	X02=25	Y02=160
3	X03=20	Y03=128
4	X04=15	Y04=102
5	X05=10	Y05=82
6	X06=5	Y06=66
Total		738

You'll get:

$E(X)$=20.6504065=$20,650.41, σ_X =8.17224837=$8,1722.48 Wrong !

The $E(X)$ you got is really $E(X|X>0)$; the σ_X you got is really $\sigma_{X|X>0}$:
$E(X|X>0)$=20.6504065=$20,650.41
$\sigma_{X|X>0}$ =8.17224837=$8,1722.48

Q6

X is the loss random variable with density function $f(x)=\frac{1}{4}e^{-\frac{x}{4}}$.
Z is the portion of the loss not covered by the insurance. Z is equal to 1 with probability of 0.4 and equal to 0 with probability of 0.6.

X and Z are independent.

Calculate $Var(XZ)$.

A 7.24 B 8.24 C 9.24 D 10.24 E 11.24

Solution D

Notice that X is exponential random variable with mean 4. Its mean is 4 and variance 16. In addition, $E(Z)=1(0.4)=0.4$; $E(Z^2)=1^2(0.4)=0.4$

$$Var(XZ)=E\left[(XZ)^2\right]-\left[E(XZ)\right]^2$$

$E\left[(XZ)^2\right]=E\left(X^2Z^2\right)$. X and Z are independent. Then X^2 and Z^2 are independent.

$$E\left[(XZ)^2\right]=E\left(X^2Z^2\right)=E\left(X^2\right)E\left(Z^2\right)$$

$$E\left(X^2\right)=E^2(X)+Var(X)=4^2+4^2=32, \quad E\left(Z^2\right)=0.4\left(1^2\right)=0.4$$

$$E\left[(XZ)^2\right]=E\left(X^2\right)E\left(Z^2\right)=32(0.4)=12.8$$

$$E(XZ)=E(X)E(Z)=4(0.4)=1.6$$

$$Var(XZ)=E\left[(XZ)^2\right]-\left[E(XZ)\right]^2=12.8-1.6^2=10.24$$

Q7 MGF for a loss random variable X is $M_x(t)=\dfrac{0.2}{0.2-t}$. The payment is $Y=80\%X+10$. What's the MGF for the payment?

A $\dfrac{1}{1-4t}e^{10t}$ B $\dfrac{0.2}{0.2-t}e^{10t}$ C $\dfrac{0.2}{0.2-t}e^{0.8t}$ D $\dfrac{1}{1-4t}e^{0.8t}$ E $\dfrac{1}{1-0.2t}e^{10t}$

Solution A

$$M_{aX+b}(t)=M_x(at)e^{bt}$$

$$M_{0.8X+10}(t)=M_X(0.8t)e^{10t}=\frac{0.2}{0.2-0.8t}e^{10t}=\frac{1}{1-4t}e^{10t}$$

Q8
The joint pdf is f(x,y)=2|x|y where -1<x<1 and 0<y<1. Calculate $E\left(X^2Y\right)$.

A 0.0 B 0.25 C 0.33 D 0.50 E 0.75

Solution C

E(X²Y)= $\int_{-1}^{1}\int_{0}^{1}\left(x^2y\right)2|x|ydydx=\int_{-1}^{1}2|x|x^2\int_{0}^{1}y^2dydx=\frac{2}{3}\int_{-1}^{1}|x|x^2dx$

$$=\frac{2}{3}\left(\int_{-1}^{0}x^{-3}dx+\int_{0}^{1}x^3dx\right)=\frac{2}{3}\left(\frac{1}{4}+\frac{1}{4}\right)=\frac{1}{3}$$

Q9
April 15 is approaching. Two taxpayers, Adam and Bob, plan to visit the same IRS office in town to ask for tax questions. Adam's arrival time to the IRS office is uniformly distributed over [10:00 am, 12:30 pm]. Bob's arrival time to the IRS office is uniformly distributed over [11:00 am, 1:00 pm]. Adam's arrival time to the IRS office is independent from Bob's arrival time to the IRS office.

The IRS office hours are from 8:00am to 4:00 pm except for a 30-minute lunch break, which lasts from 11:30 am to noon.

X = The probability that both Adam and Bob have to wait for no more than 20 minutes before they can approach an IRS clerk to ask for a question.

Y = The probability that both Adam and Bob have to wait for no more than 20 minutes before they can approach an IRS clerk to ask for a question given they are already waiting.

Calculate X + Y.

A 0.3 B 0.5 C 0.8 D 1.0 E 1.3

Solution E

X=P(Adam's waiting time $\leq$ 20 minutes)* P(Bob's waiting time $\leq$ 20 minutes)
= [1- P(Adam's waiting time>20 minutes)] * [1- P(Bob's waiting time>20 minutes)]

The only way for Adam or Bob to wait for more than 20 minutes is when they arrive between [11:30am, 11:40am]. If they arrive by 11:30 am (IRS lunch break), they will be served immediately without waiting. If they arrive shortly after 11:40am (11:45am for example), their waiting time is less than 20 minutes.

P(Adam's waiting time>20 minutes) =P(Adam arrives between 11:30am and 11:40 am)

Adam's arrival time is uniformly distributed over [10:00 am, 12:30 pm] or 2.5-hour interval, then

P(Adam arrives between 11:30am and 11:40 am)
= 10 minutes / 2.5 hours = 10 minutes / (2.5*60 minutes)=1/15

Similarly, Adam's arrival time is uniformly distributed over [11:00 am, 1 pm].

P(Bob arrives between 11:30am and 11:40 am)
= 10 minutes / 2 hours = 10 minutes / (2*60 minutes)=1/12

X= [1- P(Adam's waiting time>20 minutes)] * [1- P(Bob's waiting time>20 minutes)]
= (1- 1/15)*(1-1/12)=0.856

Next, let's calculate Y. Here the starting point is that Adam and Bob are both waiting now. To have them both wait, they must each arrive during the interval [11:30, noon]. So instead of considering the original arrival time [10:00 am, 12:30 pm] for Adam and [11:00 am, 1:00 pm] for Bob, we'll consider the updated arrival interval [11:30, noon] for both Adam and Bob. This is called shrinking the sample space (the arrival of new information reduces your original sample space).

So finding Y is reduced to:
Adam's arrival time is uniformly distributed over [11:30, noon].
Bob's arrival time is uniformly distributed over [11:30, noon].
Adam and Bob are independent.
What's the probability that Adam and Bob each arrive during [11:40 am, noon] ?

The probability that Adam arrives during [11:40 am, noon] given that he can arrive during [11:30 am, noon] is: 20 minutes / 30 minutes = 2/3.

The probability that Adam arrives during [11:40 am, noon] given that he can arrive during [11:30 am, noon] is: 20 minutes / 30 minutes = 2/3.

The probability that they have to wait at most 20 minutes given that they are already waiting is: Y=(2/3)*(2/3)=4/9=0.44

X+Y=0.856+0.44 = 1.2956

Q10

$f(x,y)=\dfrac{5(y-x)^4(y-2)(4-y)}{4y^4}$ where $0<x<y$ and $2<y<4$.

Calculate $E(X)$

A 0.00 B 0.11 C 0.28 D 0.51 E 0.75

Solution **D**

Lot of candidates will be scared by this problem. They will randomly choose an answer in the exam and move on to the next problem. This is a good strategy. You might want to do it too. However, this problem isn't hard if you know how to manipulate double integration.

Using the formula $E(X)=\int xf(x)dx$ is hard because you have to know $f(x)$. A fast approach is to use the formula $E(X)=\int\int xf(x,y)dxdy$.

$$E(X)=\int_2^4\int_0^y xf(x,y)dxdy=\int_2^4\int_0^y x\frac{5(y-x)^4(y-2)(4-y)}{4y^4}dxdy$$

$$=\int_2^4\frac{5(y-2)(4-y)}{4y^4}\int_0^y x(y-x)^4 dxdy=\int_2^4\frac{5(y-2)(4-y)}{4y^4}\left[\int_0^y x(y-x)^4 dx\right]dy$$

$$\int_0^y x(y-x)^4 dx=\frac{1}{30}y^6$$

$$E(X)=\int_2^4\frac{5(y-2)(4-y)}{4y^4}\times\frac{1}{30}y^6dy=\int_2^4\frac{(y-2)(4-y)}{24}y^2dy=\frac{45}{23}\approx 0.511$$

Q11, Q12, Q13, Q14

The liability that results from a car accident falls into two categories: the property damage liability and the personal injury liability. Let random variables X and Y represent the dollar amount (in \$10,000) of the property damage liability and the personal injury liability respectively. The pdf of X is:

$f_X(x)=\dfrac{1}{8}x$, where $0\le x\le k$ and k is a constant

Given $X = x$, Y is uniformly distributed over $[x, 2x]$.

An insurance policy is written to cover the sum of X and Y. There's a deductible of $30,000 and security loading of 30%. Calculate

Q11 The probability of having an accident where the property damage liability exceeds $20,000 and the personal injury liability exceeds $30,000.
A 0.688 B 0.712 C 0.733 D 0.752 E 0.801

Q12 The probability that insurer doesn't incur any claims.
A 0.078 B 0.082 C 0.094 D 0.102 E 0.124

Q13 The gross premium
A 4,000 B 4,300 C 4,600 D 4,900 E 5,200

Q14 The expected non-zero claim payment
A 3,250 B 3,750 C 3,985 D 4,000 E 4,150

Solution

Q11 A
Q12 C
Q13 D
Q14 E

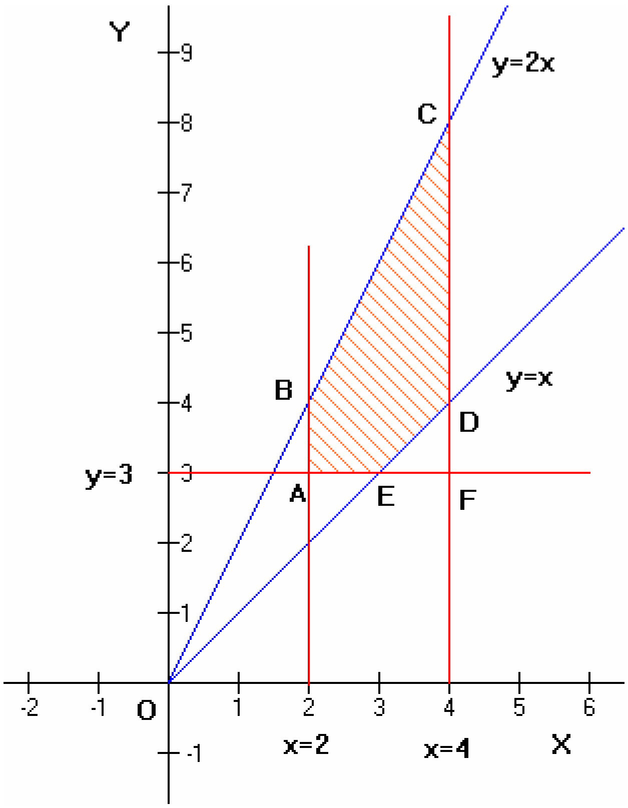
First, we need to find k. The total probability should be one $\int_0^k \frac{1}{8} x dx = 1$

$$\Rightarrow \int_0^k \frac{1}{8} x dx = \frac{1}{8}\left[\frac{1}{2}x^2\right]_0^k = \frac{k^2}{16} = 1, \quad k = 4. \text{ So } f_X(x) = \frac{1}{8}x, \text{ where } 0 \le x \le 4.$$

$$\Rightarrow f_{X,Y}(x, y) = f(x) f(y|x) = \frac{1}{8}x\left(\frac{1}{x}\right) = \frac{1}{8}, \text{ where } 0 < x \le 4 \text{ and } x \le y \le 2x.$$

The probability of having an accident where the property damage liability exceeds $20,000 and the personal injury liability exceeds $30,000 is:

$$P(X > 2 \cap Y > 3)$$



Area COD is $0 < x \le 4$ and $x \le y \le 2x$

Area ABCDE is where $0 < x \le 4$, $x \le y \le 2x$, $x > 2$, and $y > 3$.

Coordinates: A(2,3), B(2,4), C(4,8), D(4,4), F(4,4), E(3,3)

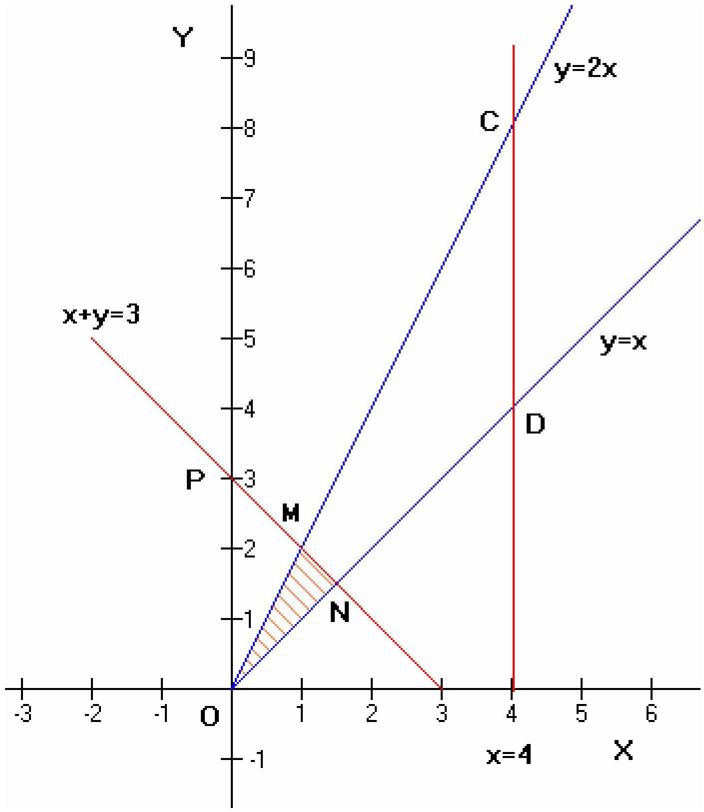
$$P(X > 2 \cap Y > 3) = \iint_{ABCDE} f(x, y)\,dxdy = \iint_{ABCDE} \frac{1}{8}dxdy = \frac{1}{8}\iint_{ABCDE} dxdy = \frac{Area\ ABCDE}{8}$$

Area ABCDE = Area ABCF – Area DEF

Area ABCF=0.5*(AB+CF)*AF=0.5*(1+5)*2=6,   Area DEF=0.5*EF*DF=0.5*1*1=0.5

Area ABCDE = 6 – 0.5 =5.5

$$\Rightarrow \quad P(X > 2 \cap Y > 3) = \frac{5.5}{8} = 0.6875$$



The shaded area is where $x + y \le 3$, $0 < x \le 4$, and $x \le y \le 2x$
In the shaded area, the total loss doesn't exceed the deductible.
Coordinates: M(1, 2) and N(1.5, 1.5)

The probability that the insurer doesn't incur any claims is:

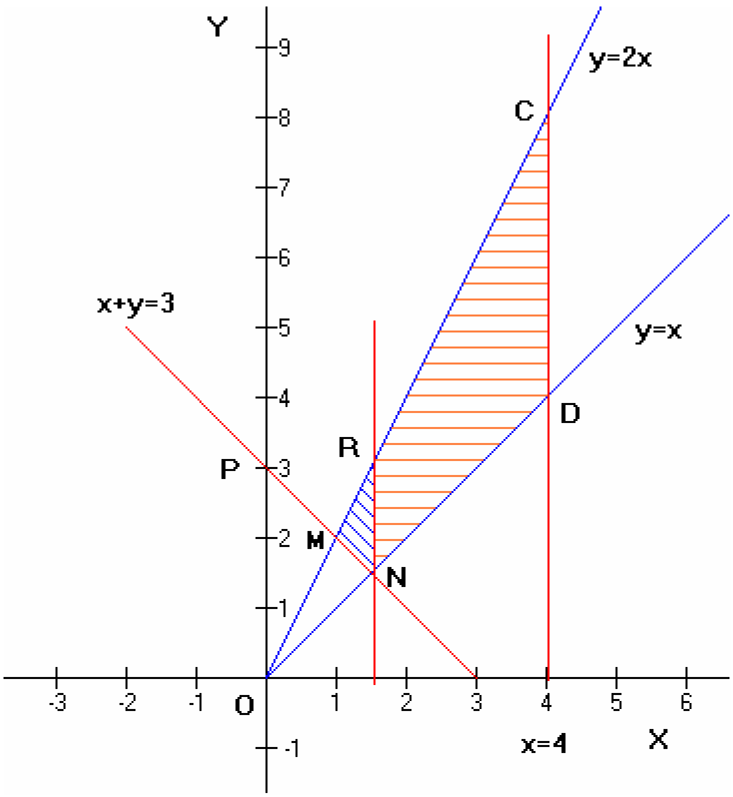
$$P(X + Y \le 3) = \iint_{MON} f(x, y)\,dxdy = \iint_{MON} \frac{1}{8}dxdy = \frac{Area\ MON}{8}$$

To find Area MON, we need to find the Coordinates of M and N. Solving the following equations:

$x + y = 3$ and $y = 2x$ $\Rightarrow$ $x = 1$ and $y = 2$. So M(1, 2)
$x + y = 3$ and $y = x$ $\Rightarrow$ $x = y = 1.5$. So N(1.5, 1.5)

Area MON = Area NOP – Area MOP
Area NOP = 0.5*OP*(X coordinate of N) = 0.5*3*1.5
Area MOP = 0.5*OP*(X coordinate of M) = 0.5*3*1
Area MON = 0.5*3*(1-0.5)=0.75

$$\Rightarrow \quad P(X + Y \le 3) = \frac{0.75}{8} = 0.09375$$



The shaded area is where the insurer has to pay claims $x+y-3$.
Coordinates: M(1, 2) and N(1.5, 1.5)

Let Z represent the claim payment. Then $Z=\max\left(X+Y-3,0\right)$. The net premium is the expected payment.

$$E\left(Z\right)=\iint_{MCDN}\left(x+y-3\right)f\left(x,y\right)dxdy=\iint_{MRN}\left(x+y-3\right)f\left(x,y\right)dxdy$$

$$+\iint_{RCDN}\left(x+y-3\right)f\left(x,y\right)dxdy$$

$$\iint_{MRN}\left(x+y-3\right)f\left(x,y\right)dxdy=\int_{1}^{1.5}\int_{3-x}^{2x}\left(x+y-3\right)\frac{1}{8}dydx=0.023438$$

$$\iint_{RCDN}\left(x+y-3\right)f\left(x,y\right)dxdy=\int_{1.5}^{4}\int_{x}^{2x}\left(x+y-3\right)\frac{1}{8}dydx=3.737$$

$E\left(Z\right)=0.023438+3.737=3.760438$ =\$3,760.44. This is the net premium.

The gross premium is: $3,760.44*(1+30%)=$4,888.57

The expected non-zero claim payment (the expected claim payment given there's a claim):

$$E(Z|Z>0)=\frac{E(Z)}{P(Z>0)}=\frac{E(Z)}{P(X+Y>3)}=\frac{E(Z)}{1-P(X+Y\le 3)}=\frac{3,760.44}{1-0.09375}=4,149.45$$

To see why $E(Z|Z>0)=\dfrac{E(Z)}{P(Z>0)}$ hold, please note that the unconditional mean is

$E(Z)=\int z\,f(z)\,dz$. To find the conditional mean, we just need to change the

unconditional density $f(z)$ to the conditional density $\dfrac{f(z)}{P(Z>0)}$. The conditional mean

is $E(Z|Z>0)=\int z\dfrac{f(z)}{P(Z>0)}dz=\dfrac{1}{P(Z>0)}\int z f(z)\,dz=\dfrac{E(Z)}{P(Z>0)}$

Q15
You take a bus to work every workday. Your journey to work consists of 3 independent components:

- The time for you to walk from home to the bus stop nearby is normally distributed with mean of 5 minutes and standard deviation of 2 minutes.

- Your bus arrives immediately after you get to the bus stop. The time for a bus to take you to the stop near your office is normally distributed with mean of 20 minutes and 6 minutes.

- The time for you to walk to work from the bus stop to your company is normally distributed with mean of 3 minutes and standard deviation of 1 minutes.

Your boss drives to work every workday. The time for him to drive to work is normally distributed with mean of 26 minutes with standard deviation of 5 minutes.

Calculate the probability that in a workday you arrive at work at least 5 minutes earlier than you boss does (assuming you leave your home and your boss leaves his home at the same time).

A 13% B 19% C 25% D 31% E 37%

Solution B

Let X represent the # of minutes it takes you to work. Let Y represent the # of minutes it takes your boss to work. Let $W = Y - X$. We are asked to calculate $P(W \geq 5)$.

Let X_1, X_2, and X_3 represent the 3 components of X. $X = X_1 + X_2 + X_3$. Because X_1, X_2, and X_3 are independent normal random variables, X is also normal.

$$E(X) = E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) = 5 + 20 + 3 = 28$$
$$Var(X) = Var(X_1 + X_2 + X_3) = Var(X_1) + Var(X_2) + Var(X_3) = 2^2 + 6^2 + 1^2 = 41$$

W is normally distributed.

$$E(W) = E(Y) - E(X) = 26 - 28 = -2$$
$$Var(W) = Var(Y - X) = Var(Y) + Var(X) = 5^2 + 41 = 66$$

$$P(W \geq 5) = 1 - P(W \leq 5) = 1 - \Phi\left[\frac{5-(-2)}{\sqrt{66}}\right] = 1 - \Phi(0.86) = 19\%$$

Q16

A manufacturing plant purchases a special product defect insurance policy. The insurance provides a payment of $1,000 if the number of defects is 20. Then for each full 5 incremental defects, the insurance pays an additional $500. However, the payment by the insurance policy can be no more than $2,800 regardless of the number of defects.

The probability is 0.19 that the plant will have less than 5 defects.
The probability is 0.36 that the plant will have less than 10 defects.
The probability is 0.51 that the plant will have less than 15 defects.
The probability is 0.64 that the plant will have less than 20 defects.
The probability is 0.75 that the plant will have less than 25 defects.
The probability is 0.84 that the plant will have less than 30 defects.
The probability is 0.91 that the plant will have less than 35 defects.
The probability is 0.96 that the plant will have less than 40 defects.
The probability is 0.99 that the plant will have less than 45 defects.
The probability is 1.00 that the plant will have less than 50 defects.

Let Y represent the payment by the insurance policy. Calculate $\dfrac{E(Y)}{\sigma(Y)}$, the ratio of the mean to the standard deviation of Y.

A 0.48 B 0.53 C 0.58 D 0.63 E 0.68

Solution E

Let X represent the # of defects.

We set $1,000 as one unit of payment to speed up calculation. This won't affect the ratio $E(Y)\ /\ \sigma(Y)$.

X>=	X<	Cumulative probability	Incremental Probability P(Y)	Raw payout (before max payment is applied)	Payout Y	Scale up probability 100*P(Y)
0	20	0.64	0.64	0	**0**	**64**
20	25	0.75	0.11	1	**1**	**11**
25	30	0.84	0.09	1.5	**1.5**	**9**
30	35	0.91	0.07	2	**2**	**7**
35	40	0.96	0.05	2.5	**2.5**	**5**
40	45	0.99	0.03	3	**2.8**	**3**
45	50	1.00	0.01	3.5	**2.8**	**1**
Total			1.00			100

In BA II Plus 1-V Statistics Worksheet, enter:
X01=0, Y01=64
X02=1, Y02=11
X03=1.5, Y03=9
X04=2, Y04=7
X05=2.5, Y05=5
X06=2.8, Y06=3+1=4

You should get: Mean $E(Y)$=0.622, σ_Y =0.91198465.
$E(Y)/\ \sigma_Y$ = 0.622 / 0.91198465 = 0.682029

Note (1) The following information is not needed to solve the problem:
The probability is 0.19 that the plant will have less than 5 defects.
The probability is 0.36 that the plant will have less than 10 defects.
The probability is 0.51 that the plant will have less than 15 defects.

Note (2) If you have time to waste in the exam or if you want to make some calculation errors, you can also calculate $E(Y)$ and σ_Y using the following formulas:

$$E(Y)=\sum y\,p(y)\,,\quad E\left(Y^2\right)=\sum y^2\,p(y)\,,\quad \sigma_Y=\sqrt{E\left(Y^2\right)-E^2(Y)}$$

Note (3) Instead of entering X06=0, Y06=3+1=4, you can also enter:
X06=0, Y06=3
X07=0, Y07=1

Note (4) Don't forget to enter X01=0, Y01=64. If you forget to enter the zero value of the random variable and the associated probability, your result will be wrong.

When using BA II Plus 1-V Statistics Worksheet, you must enter the zero value of the random variable X and the associated probability. If you don't enter $x=0$, you'll calculate $E\left(X \mid X \neq 0\right)$ and $Var\left(X \mid X \neq 0\right)$, instead of $E\left(X\right)$ and $Var\left(X\right)$.

Q17 X is a Poisson random variable. $\dfrac{F(2)}{F(1)}$ = 2.125. Find $E\left(X\right)$

A 3 B 4 C 5 D 6 E 7

Solution A

$$\frac{F(2)}{F(1)}=\frac{e^{-\lambda}\left(1+\lambda+0.5\lambda^2\right)}{e^{-\lambda}\left(1+\lambda\right)}=1+\frac{0.5\lambda^2}{1+\lambda}=2.125.$$

$\lambda^2-2.25\lambda-2.25=0$, $(\lambda-3)(\lambda+0.75)=0$, $\lambda=3$

$E\left(X\right)=\lambda=3$

Q18 Random variables X and Y are jointly uniformly distributed over the area $0 \leq |X|+|Y| \leq 1$. Calculate σ_{XY}, the standard deviation of XY.

A 0.000 B 0.016 C 0.032 D 0.105 E 0.250

Solution D

To solve this problem right, you'll need to correctly identify the 2-D plane for $0 \leq |X|+|Y| \leq 1$.

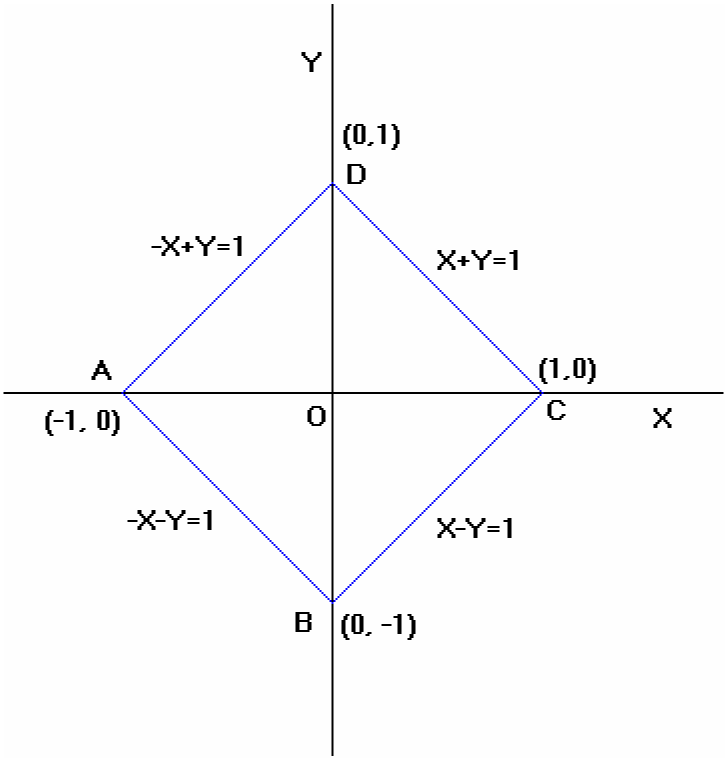
If $X \geq 0$ and $Y \geq 0$, then $|X|+|Y|=X+Y$; the 2-D plane is $0 \leq X+Y \leq 1$;

If $X \geq 0$ and $Y \leq 0$, then $|X|+|Y|=X-Y$ the 2-D plane is $0 \leq X-Y \leq 1$;

If $X \leq 0$ and $Y \leq 0$, then $|X|+|Y|=-X-Y$ the 2-D plane is $0 \leq -X-Y \leq 1$ or $-1 \leq X+Y \leq 0$;

If $X \le 0$ and $Y \ge 0$, then $|X|+|Y| = -X+Y$ the 2-D plane is $0 \le -X+Y \le 1$ or $-1 \le X-Y \le 0$.

So the 2-D plane is ABCD (you should verify this).



We'll need to calculate $f(x,y)=k$. The total probability should be one:

$$\iint_{ABCD} f(x,y)\,dydx = \iint_{ABCD} k\,dydx = k \iint_{ABCD} dydx = k \times (Area\ ABCD) = 1$$

$$Area\ ABCD = 4 \times Area\ AOB = 4 \times \frac{1}{2} = 2 \qquad \Rightarrow f(x,y) = k = \frac{1}{2}$$

$$Var(XY) = E(X^2Y^2) - E^2(XY)$$

$$E(X^2Y^2) = \iint_{ABCD} x^2y^2 f(x,y)\,dydx = \frac{1}{2} \iint_{ABCD} x^2y^2\,dydx$$

x^2y^2 is symmetric inside AOB, AOD, COB, and COD.

$$\Rightarrow \quad \iint_{ABCD} x^2 y^2 \, dydx = 4 \iint_{COD} x^2 y^2 \, dydx .$$

$$\iint_{COD} x^2 y^2 \, dydx = \int_0^1 \int_0^{1-x} x^2 y^2 \, dydx = \int_0^1 x^2 \left(\int_0^{1-x} y^2 \, dy \right) dx = \frac{1}{3} \int_0^1 x^2 (1-x)^3 dx$$

To quickly solve this integration, you'll want to use the shortcut developed in the chapter on beta distribution (you should memorize this integration shortcut):

$$\int_0^1 p^m (1-p)^n \, dp = \frac{1}{(m+n+1) C_{m+n}^m}$$

$$\Rightarrow \quad \iint_{COD} x^2 y^2 \, dydx = \frac{1}{3} \int_0^1 x^2 (1-x)^3 dx = \frac{1}{3} \frac{1}{(2+3+1) C_{2+3}^2} = \frac{1}{3} \times \frac{1}{(6) C_5^2} = \frac{1}{180}$$

$$\Rightarrow \quad \iint_{ABCD} x^2 y^2 \, dydx = 4 \iint_{COD} x^2 y^2 \, dydx = 4 \times \frac{1}{180}$$

$$\Rightarrow \quad E\left(X^2 Y^2 \right) = \frac{1}{2} \iint_{ABCD} x^2 y^2 \, dydx = \frac{1}{2} \times 4 \times \frac{1}{180} = \frac{1}{90}$$

Next, we need to calculate $E^2(X\,Y)$. Please note that $X\,Y$ is positive inside AOB and COD and negative inside AOD and COB. Consequently,

$$E^2(X\,Y) = \iint_{ABCD} x\,y\, f(x,y) \, dydx = \frac{1}{2} \iint_{ABCD} x\,y\, dydx$$

$$= \frac{1}{2} \left[\iint_{AOB} x\,y\, dydx + \iint_{AOD} x\,y\, dydx + \iint_{COD} x\,y\, dydx + \iint_{COB} x\,y dydx \right] = 0$$

$$\Rightarrow \quad Var(X\,Y) = E\left(X^2 Y^2 \right) - E^2(X\,Y) = E\left(X^2 Y^2 \right) = \frac{1}{90}$$

$$\sigma_{X\,Y} = \sqrt{Var(X\,Y)} = \sqrt{\frac{1}{90}} = 0.1054$$

By the way, if the problem asks you to find $Var(X)$, you can find the answer using the approach above:

$$Var(X) = E\left(X^2 \right) - E^2(X)$$

$$E\left(X^2\right)=\iint_{ABCD} x^2 f\left(x,y\right)dydx=\frac{1}{2}\iint_{ABCD} x^2 dydx$$

x^2 is symmetric inside AOB, AOD, COB, and COD.

$$\Rightarrow \quad \iint_{ABCD} x^2\,dydx=4\iint_{COD} x^2\,dydx.$$

$$\iint_{COD} x^2 dydx=\int_0^1\int_0^{1-x} x^2\,dydx=\int_0^1 x^2\left(\int_0^{1-x} dy\right)dx=\int_0^1 x^2\left(1-x\right)dx=\frac{1}{12}$$

$$\Rightarrow \quad \iint_{ABCD} x^2\,dydx=4\iint_{COD} x^2 dydx=4\times\frac{1}{12}=\frac{1}{3}$$

$$\Rightarrow \quad E\left(X^2\right)=\frac{1}{2}\iint_{ABCD} x^2\,dydx=\frac{1}{2}\times\frac{1}{3}=\frac{1}{6}$$

X is positive inside AOB and COD and negative inside AOD and COB. Consequently, $E\left(X\right)=0$

$$\Rightarrow \quad Var\left(X\right)=E\left(X^2\right)-E^2\left(X\right)=E\left(X^2\right)=\frac{1}{6}$$

$$\Rightarrow \quad \sigma_X=\frac{1}{\sqrt{6}}=0.408$$

Q19

Random variables X,Y,Z have the following joint pdf: f(x,y,z)=$\dfrac{1}{\sqrt{xy}}$ where 0<x<y<z<1.

Calculate the probability that at least two of the three random variables X,Y,Z are less than 0.1.

A 0.20 B 0.25 C 0.30 D 0.35 E 0.40

Solution A

"At least two of the three random variables X,Y,Z are less than 0.1" is the same as Y<0.1. If Y<0.1, then we are guaranteed to have X<Y<0.1 regardless of whether Z<0.1 or Z≥0.1.

y<0.1 $\Rightarrow$ 0<x<y ∩ y<z<1 ∩ 0<y<0.1

$$P(Y<0.1)=\int_{y=0}^{0.1}\int_{z=y}^{1}\int_{x=0}^{y} x^{-0.5}y^{-0.5}dxdzdy=\int_{y=0}^{0.1} y^{-0.5}\int_{z=y}^{1}\int_{x=0}^{y} x^{-0.5}dxdzdy$$

$$=\int_{y=0}^{0.1} y^{-0.5}\int_{z=y}^{1} 2y^{0.5}dzdy=2\int_{y=0}^{0.1}\int_{z=y}^{1} dzdy=2\int_{y=0}^{0.1}(1-y)\,dy$$

$$=-\left[(1-y)^2\right]_0^{0.1}=1-0.9^2=0.19$$

If the problem asks you to find the probability that at least two of the three random variables X,Y,Z are less than a where $0<a<1$, then

$P(Y<a)=-\left[(1-y)^2\right]_0^a=1-(1-a)^2=a(2-a)$

For example, $P(Y<0.25)=0.25(2-0.25)=0.4375$

Q20 In a small town, 55% are men and 45% are women. 74% of the women read the local newspaper everyday. Given that someone reads the local newspaper everyday, the probability that the reader is a male is 57.53%. Calculate the percentage of the men who read the local newspaper everyday.

A 0.78 B 0.82 C 0.88 D 0.92 E 0.98

Solution B

Event: Someone reads the local newspaper everyday.

Segment	Segment size	Segment's probability to produce the event	Segment's contribution amount	Segment's contribution %
Men	55%	x	0.55 x	$\frac{0.55x}{0.55x+0.333}=57.53\%$
Women	45%	74%	0.45(0.74)=0.333	$\frac{0.333}{0.55x+0.333}=1-57.53\%=42.27\%$
Total	100%		0.55 x +0.333	

Solving $\frac{0.55x}{0.55x+0.333}=57.53\%$, we get: x =82%

If you prefer the formula driven approach, this is how:

$$P(men|read)=\frac{P(men\cap read)}{P(read)}$$

$$P(read)=P(men)P(read|men)+P(women)P(read|women)$$
$$P(men\cap read)=P(men)P(read|men)$$

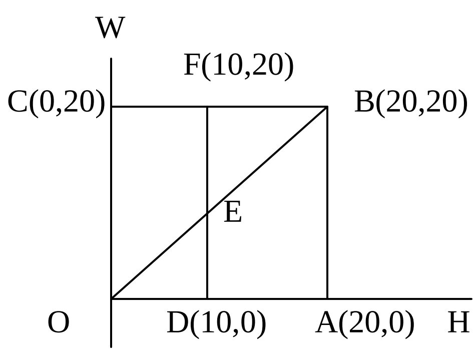
$$\Rightarrow\quad P(men|read)=\frac{P(men)P(read|men)}{P(men)P(read|men)+P(women)P(read|women)}$$

$$57.53\%=\frac{0.55P(read|men)}{0.55P(read|men)+0.45(0.74)},\quad P(read|men)=82\%$$

Q21
A wife and husband's lives are uniformly distributed on (0,20) in years. Find the conditional probability that the husband outlives the wife given that the husband is still alive 10 years from today.

A 0.55 B 0.65 C 0.75 D 0.85 E 0.95

Solution **C**



$$P(H>W\mid H>10)=\frac{P(H>W\cap H>10)}{P(H>10)}=\frac{Area\ ABED}{Area\ ABFD}$$

$$=\frac{0.5(10+20)(10)}{10(20)}=0.75$$

Q22
X is exponentially distributed with mean of 1. $Y = |X - 4|$. Calculate $f_Y(2)$, the pdf of Y at $y = 2$.

A 0.14 B 0.17 C 0.21 D 0.25 E 0.31

Solution A

$$F_X(x) = 1 - e^{-x}$$
$$F_Y(y) = P(Y \le y) = P(|X-4| \le y) = P(-y \le X - 4 \le y) \backslash$$
$$= P(-y+4 \le X \le y+4) = F_X(y+4) - F_X(-y+4)$$
$$= 1 - e^{-(y+4)} - 1 + e^{-(-y+4)} = e^{y-4} - e^{-(y+4)}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}\left[e^{y-4} - e^{-(y+4)}\right] = e^{y-4} + e^{-(y+4)}$$

$$f_Y(2) = e^{2-4} + e^{-(2+4)} = e^{-2} + e^{-6} = 0.14$$

Q23
Random variable X has the following moment generating function:

$$M_X(t) = (1-2t)^{-\frac{1}{2}}$$

Calculate the coefficient of variation.

A 0.2 B 0.5 C 0.8 D 1.1 E 1.4

Solution E

$$\frac{d}{dt} M_X(t) = \frac{d}{dt}\left[(1-2t)^{-\frac{1}{2}}\right] = (1-2t)^{-\frac{3}{2}},$$
$$E(X) = \left[\frac{d}{dt} M_X(t)\right]_{t=0} = (1-2\times 0)^{-\frac{3}{2}} = 1$$

$$\frac{d^2}{dt^2} M_X(t) = \frac{d}{dt}\left[(1-2t)^{-\frac{3}{2}}\right] = 3(1-2t)^{-\frac{5}{2}}$$

$$E\left(X^2\right)=\left[\frac{d^2}{dt^2}M_X\left(t\right)\right]_{t=0}=3\left(1-2\times0\right)^{-\frac{5}{2}}=3$$

$$Var\left(X\right)=E\left(X^2\right)-E^2\left(X\right)=3-1^2=2$$

Coefficient of variation: $\dfrac{\sigma_X}{E\left(X\right)}=\dfrac{\sqrt{2}}{1}=1.41$

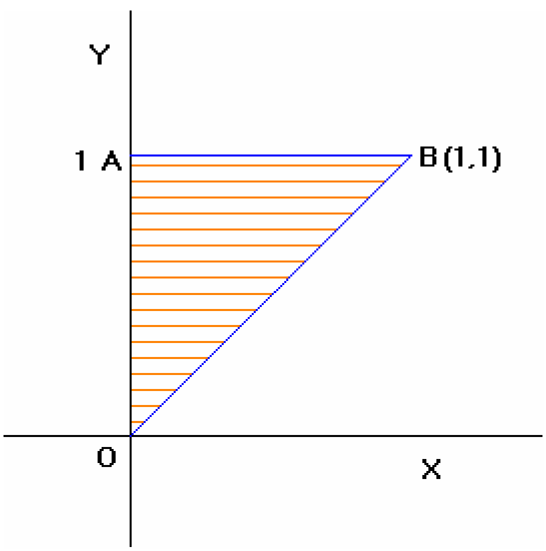
Q24
Random loss variables X and Y have the following joint pdf:

$f\left(x,y\right)=x+k\,y^2$, where $0\le x\le y\le 1$

An insurance policy is written to cover $X+Y$. The maximum claim payment is \$1.
Calculate the net premium

A 0.\$75 B 0.\$94 C \$1.0 D \$1.6 E \$2.2

Solution B

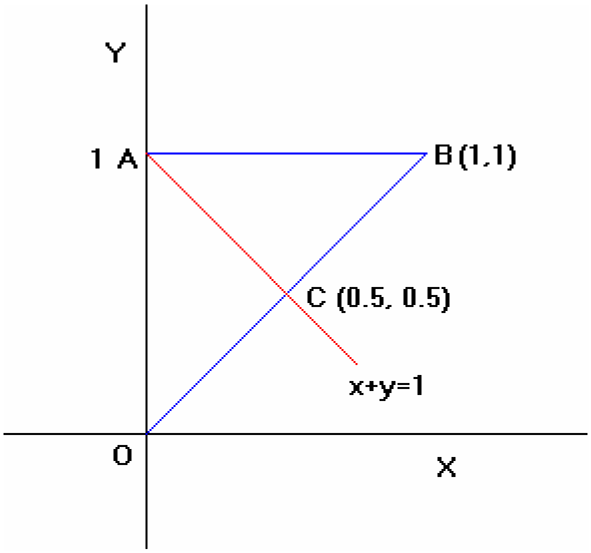


The shaded area AOB is $0\le x\le y\le 1$

First, we need to calculate k. $\displaystyle\iint_{AOB} f\left(x,y\right)dxdy=1$

$$\iint_{AOB} f(x,y)\,dxdy = \int_0^1 \int_0^y f(x,y)dxdy = \int_0^1 \int_0^y \left(x + k\,y^2\right)dxdy = \int_0^1 \left[\frac{1}{2}x^2 + kxy\right]_0^y dy$$

$$= \int_0^1 \left(\frac{1}{2}y^2 + ky^3\right)dy = \left[\frac{1}{6}y^3 + \frac{k}{4}y^4\right]_0^1 = \frac{1}{6} + \frac{k}{4} = 1, \quad k = \frac{10}{3}$$



The insurer pays $x + y$ in AOC. It pays \$1 at ABC.

The expected claim is: $\iint_{AOC} (x+y)\,f(x,y)\,dxdy + \iint_{ABC} f(x,y)\,dxdy$

$$\iint_{AOC} (x+y)\,f(x,y)\,dxdy = \int_0^{0.5} \int_x^{1-x} (x+y)\left(x + \frac{10}{3}y^2\right)dydx = 0.22569$$

Two ways to calculate $\iint_{ABC} f(x,y)\,dxdy$:

$$\iint_{ABC} f(x,y)\,dxdy = \int_{0.5}^{1} \int_{1-y}^{y} (x+y)\left(x + \frac{10}{3}y^2\right)dxdy = 0.71528$$

Or $\iint_{ABC} f(x,y)\,dxdy = 1 - \iint_{AOC} f(x,y)\,dxdy = 1 - \iint_{AOC} (x+y)\,f(x,y)\,dxdy$

$$= 1 - \int_0^{0.5} \int_x^{1-x} \left(x + \frac{10}{3}y^2\right)dydx = 1 - 0.28472 = 0.71528$$

$$\iint_{AOC} (x+y) f(x,y) dxdy + \iint_{ABC} f(x,y) dxdy = 0.22569 + 0.71528 = 0.94097$$

So the net premium is $0.94.

Q25
A device has 4 duplicate components working in parallel. The device works as long as at least one component works; it fails only if all four components fail simultaneously. The time-to-failure of the 4 components are independent exponential random variables with mean of 1, 2, 3, and 4 hours respectively.

Calculate the probability that the device is still working 5 hours later.

A 0.32 B 0.37 C 0.42 D 0.47 E 0.52

Solution D

Let X_1, X_2, X_3, and X_4 represent the time-to-failure of the 4 components. Let Y represent the device's time-to-failure. Then

$$Y = \max(X_1, X_2, X_3, X_4)$$

$$P(Y>5) = P[\max(X_1, X_2, X_3, X_4) > 5] = 1 - P[\max(X_1, X_2, X_3, X_4) \le 5]$$

If you write $P(Y \ge 5)$, that's OK. $P(Y \ge 5) = P(Y > 5)$ because Y is continuous.

$$P[\max(X_1, X_2, X_3, X_4) \le 5] = P(X_1 \le 5) P(X_2 \le 5) P(X_3 \le 5) P(X_4 \le 5)$$

If X is an exponential random variable with mean of θ, then

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}},\ P(X \le x) = F(x) = 1 - e^{-\frac{x}{\theta}},$$

$$P(X_1 \le 5) P(X_2 \le 5) P(X_3 \le 5) P(X_4 \le 5)$$
$$= \left(1-e^{-\frac{5}{1}}\right)\left(1-e^{-\frac{5}{2}}\right)\left(1-e^{-\frac{5}{3}}\right)\left(1-e^{-\frac{5}{4}}\right) = 0.5276$$

$$P(Y>5) = 1 - P[\max(X_1, X_2, X_3, X_4) \le 5] = 0.4724$$

Q26
Loss random variable X has the following pdf:

$$f(x) = 0.02x\text{, where } 0 < x < 10$$

The insurer has a deductible of 4 per loss.

Calculate the expected claim payment net of deductible.

A 2.5 B 2.7 C 2.9 D 3.4 E 4.2

Solution **C**

Let Y represent the claim payment (net of deductible). Then $Y = 0$ if $X \leq 4$ and $Y = X - 4$ if $X > 4$. We are asked to find $E(Y)$

$$E(Y) = \int_0^{10} y(x) f(x) dx = \int_0^4 0 f(x) dx + \int_4^{10} (x-4) 0.02x dx = \int_4^{10} (x-4) 0.02x dx = 2.88$$

Q27
Two discrete random variables X and Y are jointly distributed over a series of points. The joint probability mass function is $p_{X,Y}(x, y) = \frac{1}{28}$. You are also given:

- X =0, 1, 2, 3, 4, 5, 6
- Y =0, …, X (i.e. for each value in X, Y is a non-negative integer ranging from 0 to X)

Calculate $\frac{Var(X)}{Var(Y)}$

A 0 B 0.5 C 1 D 1.5 E 2

Solution **C**

To find $Var(X)$, we need to list all the possible value of X and the probability mass function.

X	Y	$p_X(x)$
0	0	1/28
1	0, 1	2/28
2	0, 1, 2	3/28
3	0, 1, 2, 3	4/28
4	0, 1, 2, 3, 4	5/28
5	0, 1, 2, 3, 4, 5	6/28
6	0, 1, 2, 3, 4, 5, 6	7/28
Total		28/28=1

Then $E(X)=\sum xp_X(x)$ $\quad E(X^2)=\sum x^2 p_X(x)$ $\quad Var(X)=E(X^2)-E^2(X)$

However, in the exam, the above formulas are time-consuming and prone to errors. You should use BA II Plus 1-V Worksheet.

In BA II Plus 1-V Statistics Worksheet, enter:
X01=0, Y01=1
X02=1, Y02=2
X03=2 Y03=3
X04=3, Y04=4
X05=4, Y05=5
X06=5, Y06=6
X07=6, Y07=7

You should get: Mean $E(X)$=4, σ_X =1.73205081
$Var(X)=\sigma_X^2=3$

Similarly, you can find $Var(Y)$

X	Y	Y	$p_Y(y)$
0	0	0	7/28
1	0, 1	1	6/28
2	0, 1, 2	2	5/28
3	0, 1, 2, 3	3	4/28
4	0, 1, 2, 3, 4	4	3/28
5	0, 1, 2, 3, 4, 5	5	2/28
6	0, 1, 2, 3, 4, 5, 6	6	1/28
		Total	28/28=1

In BA II Plus 1-V Statistics Worksheet, enter:
X01=0, Y01=7
X02=1, Y02=6
X03=2 Y03=5
X04=3, Y04=4

X05=4, Y05=3
X06=5, Y06=2
X07=6, Y07=1

You should get: Mean $E(Y)$=3, σ_Y =1.73205081
$Var(Y)=\sigma_Y^2=3$

$$\Rightarrow \quad \frac{Var(X)}{Var(Y)}=1$$

Q28
A company pays a benefit of 100 for each of its 1000 employees if an employee dies next year. The probability that an employee dies next year is 2%. What is the amount the company needs to have in a fund in order to ensure a 95% chance that it can cover the loss?

A 2,500 B 2,800 C 3,100 D 3,400 E 3,700

Solution B

Let K = # of deaths next year. K is binomial random variable with $n=1000$ and $p=0.02$. K is approximately normal with $EK=np=1000(0.02)=20$ and $Var(K)=npq=1000(0.02)(0.98)=19.6$

Let $X=100K$ represent the total payment. We need to calculate the 95-th percentile of X, X_{95}. Since $X=100K$ is an increasing function of K with one-to-one mapping, 95-th percentile of X corresponds to 95-th percentile of K. So we just need to find K_{95}. Then $X_{95}=100K_{95}$.

$$\Phi\left(\frac{K_{95}-0.5-20}{\sqrt{19.6}}\right)=0.95, \qquad \frac{K_{95}-0.5-20}{\sqrt{19.6}}=1.645$$

$$K_{95}=0.5+20+1.645\sqrt{19.6}=27.783$$

$X_{95}=100K_{95}$=100*27.783=2,778

Q29
Random variables X and Y have the following joint pdf:

$f(x,y)=k\dfrac{x}{y}$, where $0<x<y<1$ and k is a constant.

Calculate $E\left(X\middle|Y=\dfrac{2}{3}\right)+Var\left(X\middle|Y=\dfrac{2}{3}\right)$

A 0.17 B 0.27 C 0.37 D 0.47 E 0.57

Solution **D**

$$f\left(x\middle|y=\frac{2}{3}\right)=\frac{f\left(x,y=\frac{2}{3}\right)}{P\left(y=\frac{2}{3}\right)}=\frac{\frac{3}{2}kx}{\int_0^{\frac{2}{3}}f\left(x,y=\frac{2}{3}\right)dx}=\frac{\frac{3}{2}kx}{\int_0^{\frac{2}{3}}\frac{3}{2}kxdx}=cx,$$

where $0<x<\dfrac{2}{3}$ and $\dfrac{\frac{3}{2}k}{P\left(y=\frac{2}{3}\right)}=\dfrac{\frac{3}{2}k}{\int_0^{\frac{2}{3}}f\left(x,y=\frac{2}{3}\right)dx}=c$ is a constant.

You don't need to worry about actually finding k or $P\left(y=\dfrac{2}{3}\right)$. You just need to find c.

$f\left(x\middle|y=\dfrac{2}{3}\right)=cx$ should add up to one over the range $0<x<\dfrac{2}{3}$:

$$\Rightarrow \quad \int_0^{\frac{2}{3}}f\left(x\middle|y=\frac{2}{3}\right)dx=\int_0^{\frac{2}{3}}cxdx=1, \quad c=4.5.$$

$$\Rightarrow \quad f\left(x\middle|y=\frac{2}{3}\right)=4.5x, \text{ where } 0<x<\frac{2}{3}.$$

So $E\left(X \middle| Y=\frac{2}{3}\right)=\int_0^{\frac{2}{3}} xf\left(x \middle| y=\frac{2}{3}\right)dx=\int_0^{\frac{2}{3}} x(4.5x)dx=\left[1.5x^3\right]_0^{\frac{2}{3}}=1.5\left(\frac{2}{3}\right)^3=\frac{4}{9}$ =0.444

$$E\left[\left(X \middle| Y=\frac{2}{3}\right)^2\right]=\int_0^{\frac{2}{3}} x^2 f\left(x \middle| y=\frac{2}{3}\right)dx=\int_0^{\frac{2}{3}} x^2(4.5x)dx=\left[\frac{4.5}{4}x^4\right]_0^{\frac{2}{3}}=\frac{4.5}{4}\left(\frac{2}{3}\right)^4=\frac{2}{9}$$

$Var\left(X \middle| Y=\frac{2}{3}\right)=\frac{2}{9}-\left(\frac{4}{9}\right)^2$ =0.02469

$E\left(X \middle| Y=\frac{2}{3}\right)+Var\left(X \middle| Y=\frac{2}{3}\right)$=0.4444 + 0.02469=0.47

Q30

Random variable X has the following moment generating function: $M_X(t)=\frac{1}{1-\theta t}$, where θ is a positive constant. Calculate the probability that X is one standard deviation from its mean.

A 0.46 B 0.56 C 0.66 D 0.76 E 0.86

Solution **E**

Notice that X is an exponential random variable with mean θ and standard deviation of θ.

$$P\left[-\sigma_X \le X-E(X)\le \sigma_X\right]=P(-\theta\le X-\theta\le\theta)=P(0\le X\le 2\theta)=F(2\theta)$$

$$=1-e^{-\frac{2\theta}{\theta}}=1-e^{-2}=0.8647$$

Final tips on taking Exam P

1. The goal of combat is to win the war, not individual battles. The goal of taking Exam P is to get a 6 so you can look for a job, not to get a 10.

2. If you need to speak in an important meeting (such as before Congress), always prepare a script well in advance and rehearse your script. When preparing for an important exam such P, always build a 3 minute solution script ahead of time for each of the tested problems in Sample P and any newly released P exams (if any). Then when taking the exam, just regurgitate the script and solve the repeatedly tested problems.

3. In the exam, if a problem is brand new (this type of problems have not been tested before), make one attempt. If you can't solve, just guess an answer, and move on to the next problem.

4. Focus on mastering the fundamentals. Difficult distributions such as negative binomial, hypergeometric, Weibull, Pareto, beta, Chi-square, and bivariate normal aren't tested in the Sample Exam P problems. Since Sample Exam P reflects the level of knowledge you need to have to pass Exam P, chances are that these difficult distributions won't show up in your exam. As a result, you can probably ignore these distributions. However, if you really can't sleep well if you ignore these distributions, you might learn some basic knowledge about these distributions. Don't over-study these difficult distributions.

5. Master Sample Exam P before taking your exam. Put yourself in the exam-like condition and practiced Sample P exam and any newly released P exams (if any). Work and rework until you can solve Sample P exam and any newly released P exams (if any) 100% right in the exam-like condition. Never walk into the exam room without mastering Sample P exam and any newly released P exams (if any).

6. For those wanting for additional practice problems, refer to http://www.actuarialoutpost.com/actuarial_discussion_forum/. You'll find many practice problems there.

7. If you failed Exam P, don't give up. Many candidates eventually passed Exam P after failing it multiple times.

About the author

Yufeng Guo was born in central China. After receiving his Bachelor's degree in physics at Zhengzhou University, he attended Beijing Law School and received his Masters of law. He was an attorney and law school lecturer in China before immigrating to the United States. He received his Masters of accounting at Indiana University. He has pursued a life actuarial career and passed exams 1, 2, 3, 4, 5, 6, and 7 in rapid succession after discovering a successful study strategy.

Mr. Guo's exam records are as follows:

Fall 2002	Passed Course 1
Spring 2003	Passed Courses 2, 3
Fall 2003	Passed Course 4
Spring 2004	Passed Course 6
Fall 2004	Passed Course 5
Spring 2005	Passed Course 7

Mr. Guo currently teaches an online prep course for Exam P, FM, and M. For more information, visit http://guo.coursehost.com.

If you have any comments or suggestions, you can contact Mr. Guo at yufeng_guo@msn.com.

Please note that if I find any errors, I will post the errata at http://guo.coursehost.com.

Value of this PDF study manual

1. Don't pay the shipping fee (can cost $5 to $10 for U.S. shipping and over $30 for international shipping). Big saving for Canadian candidates and other international exam takers.

2. Don't wait a week for the manual to arrive. You download the study manual instantly from the web and begin studying right away.

3. Load the PDF in your laptop. Study as you go. Or if you prefer a printed copy, you can print the manual yourself.

4. Use the study manual as flash cards. Click on bookmarks to choose a chapter and quiz yourself.

5. Search any topic by keywords. From the Adobe Acrobat reader toolbar, click Edit ->Search or Edit ->Find. Then type in a key word.

User review of Mr. Guo's P Manual

Mr. Guo's P manual has been used extensively by many Exam P candidates. For user reviews of Mr. Guo's P manual at http://www.actuarialoutpost.com, click here Review of the manual by Guo.

Testimonies:

"**Second time I used the Guo manual and was able to do some of the similar questions in less than 25% of the time because of knowing the shortcut**."
Testimony #1 of the manual by Guo

"I just took the exam for the second time and feel confident that I passed. I used Guo the second time around. It was very helpful and gives a lot of shortcuts that I found very valuable. I thought the manual was kind of expensive for an e-file, but if it helped me pass it was well worth the cost."
Testimony # 2 of the manual by Guo

"I took the last exam in Feb 2006, and I ran out of time and I ended up with a five. I needed to do the questions quicker and more efficiently. The Guo's study guide really did the job."
Testimony #3 of the manual by Guo

Bonus Problems

Problem 1 *A system consists of two machines that work together. The system works only if both machines work. The system fails if either machine fails. The time-until-failure of each machine is independent from each other. The joint moment generating function of each machine's time-until-failure is* $\dfrac{1}{20st-5t-4s+1}$. *Calculate the probability that the system is still working after* 1 *hour.*

Solution

Let X and Y represent the time-until-failure of the two machines. Since X and Y are independent, we have:

$M_{X,Y}(s,t) = M_X(s)\,M_Y(t)$

On the other hand, we are given:

$M_{X,Y}(s,t) = \dfrac{1}{20st-5t-4s+1} = \dfrac{1}{1-4s} \times \dfrac{1}{1-5t}$

$\dfrac{1}{1-4s}$ is the MGF of an exponential random variable with mean $\theta = 4$

$\dfrac{1}{1-5t}$ is the MGF of an exponential random variable with mean $\theta = 5$

Hence X and Y are two independent exponential random variables. One of these two exponential random variables has a mean of 4 and the other 5.

$P(X>1 \cap Y>1) = P(X>1)\,P(Y>1) = e^{-1/4}e^{-1/5} = 0.637\,63$

Problem 2 *X and Y are independent random variables. $E(X) = 0.5E(Y)$. The coefficient of variation of X is 5; the coefficient of variation of Y is 8. Find the coefficient of variation of $10(X+Y)$.*

Solution

Please note that the definition of the coefficient of the variation of a random variable Z is

$coe_Z = \dfrac{\sqrt{Var(Z)}}{E(Z)}$

It then follows that for any non zero constant m

$coe_{mZ} = \dfrac{\sqrt{Var(mZ)}}{E(mZ)} = \dfrac{\sqrt{m^2 Var(Z)}}{mE(Z)} = \dfrac{\sqrt{Var(Z)}}{E(Z)} = coe_Z$

$coe_{10(X+Y)} = coe_{X+Y} = \dfrac{\sqrt{Var(X+Y)}}{E(X+Y)}$

$coe_X = \dfrac{\sqrt{Var(X)}}{E(X)} = 5 \qquad Var(X) = 5^2E^2(X) = 5^2\left(0.5^2\right)E^2(Y)$

$coe_Y = \dfrac{\sqrt{Var(Y)}}{E(Y)} = 8 \qquad Var(Y) = 8^2E^2(Y)$

$Var(X+Y) = Var(X) + Var(Y) = \left(1+8^2\right)E^2(Y)$

$\sqrt{Var(X)+Var(Y)} = \sqrt{5^2\left(0.5^2\right)+8^2}E(Y)$

$E(X+Y) = E(Y) + E(Y) = 1.5E(Y)$

$coe_{X+Y} = \frac{\sqrt{Var(X+Y)}}{E(X+Y)} = \frac{\sqrt{5^2(0.5^2)+8^2}E(Y)}{1.5E(Y)} = \frac{\sqrt{5^2(0.5^2)+8^2}}{1.5} = 5.5877$

$coe_{10(X+Y)} = coe_{X+Y} = 5.5877$

Problem 3 *There are 15 people. 10 people from Department A and the other from Department B. The probability of having an accident is $p = 0.05$ for each person. What is the probability of at least one accident from Department A given that there are 4 accidents?*

Solution

The probability that of 15 people 4 have accidents and 11 don't is a binomial distribution: $C_{15}^4 p^4 q^{11}$

The probability that department B has 4 accidents and department A has 0 accident (i.e. all accidents are from department B) is

$C_5^4 p^4 q \times C_{10}^{10} p^0 q^{10}$

The probability that all accidents are from department B given there are 4 accidents is:

$\frac{C_5^4 p^4 q \times C_{10}^{10} p^0 q^{10}}{C_{15}^4 p^4 q^{11}} = \frac{C_5^4 C_{10}^{10}}{C_{15}^4} = \frac{4 \times 3 \times 2 \times 1}{15 \times 14 \times 13 \times 12} = \frac{1}{1365}$

The probability of at least one accident from A given that there are 4 accidents is

$1 - \frac{1}{1365} = 0.99927$