CS 70 Discrete Mathematics and Probability Theory
Fall 2015 Jean Walrand HW 13 Solutions

1. Statistical hypothesis testing
   On one of the Mythbusters episodes[1], the Mythbusters decided to run an experiment to test whether toast tends to land buttered side down.

   At the beginning of the episode, Adam and Jamie built a first attempt at a mechanical rig to drop toast in a controlled fashion. When they tested it on 10 unbuttered pieces of toast as a sanity check, 7 pieces fell upside down and 3 pieces fell right-side up. Adam concluded based upon these numbers that this first rig was obviously biased, so he threw it away in disgust and they built a new rig. Was Adam right, or is this just another case where he jumps to conclusions too quickly?

   Let p denote the probability that, if we drop 10 pieces of unbuttered toast from an unbiased rig (i.e., a rig where each unbuttered piece of toast has a 50% chance of falling upside down and a 50% chance of falling right-side up), 7 or more of the pieces of toast land the same way. In other words, p is the probability of the event that at least 7 pieces land right-side up, or at least 7 pieces land upside down, when dropping from an unbiased rig.

   (a) As a warmup, compute the exact probability that if we flip a fair coin 10 times, we see 0, 1, 2, 3, 7, 8, 9, or 10 heads.

   Answer: Let X be the number of flips that come up heads. Then $X \sim \text{Binomial}(10, 0.5)$.

$$\begin{aligned}
\Pr[(0 \le X \le 3) \vee (7 \le X \le 10)] &= \binom{10}{0}0.5^{10} + \binom{10}{1}0.5^{10} + \binom{10}{2}0.5^{10} + \binom{10}{3}0.5^{10} + \\
&\quad \binom{10}{7}0.5^{10} + \binom{10}{8}0.5^{10} + \binom{10}{9}0.5^{10} + \binom{10}{10}0.5^{10} \\
&= \frac{11}{32} \\
&= 0.34375.
\end{aligned}$$

   (b) Now, back to the Mythbusters. With p defined as above, calculate p exactly.

   Answer: p is the probability that the number of toast pieces landing right-side up is between 0 and 3 or between 7 and 10, all ranges inclusive. If we think of the toast as a coin and the outcome of right-side up as Heads, then p is exactly the probability computed in part (a). Hence, $p = 0.34375$.

   (c) Use p to decide whether the rig appears biased, using the following rules:

   - If $p > 0.05$, conclude that we cannot rule out the possibility that the rig is unbiased. The rig might be perfectly good as it is.
     (The intuition is: Oh man, that totally could've happened by chance.)

[1] Season 3, episode 4, air date: March 9, 2005.

- If $p \leq 0.05$, with 95% confidence we can conclude that the rig appears to be biased.
  (Sure, it's possible that this rule could lead us astray. Even if our calculations show $p \leq 0.05$, it's in principle *possible* that the rig is unbiased and the observations were just a big coincidence. However, this would require assuming that an event of probability 0.05 or less happened, which is by definition pretty rare. Put another way, if we conclude that the rig is biased whenever $p \leq 0.05$, then we'll wrongly throw away a perfectly good rig at most 5% of the time. This seems good enough.)

To put it another way, this decision rule gives us a way to test the hypothesis that the rig is unbiased: if $p \leq 0.05$, we reject the hypothesis (with 95% confidence), otherwise if $p > 0.05$ we are unable to reject it (at 95% confidence level).

Using your value of p and this decision rule, decide whether Adam was right to conclude that his first rig was biased, or whether he jumped to conclusions too quickly.

Answer: In part (b), we found that $p = 0.34375$. This says that the outcome of 7 right-side up and 3 right-side down pieces of toast, or 7 right-side down and 3 right-side up pieces of toast could be produced by an unbiased rig with probability 0.34375. Since this probability is larger than 0.05, the observed data doesn't give us enough evidence to reject the hypothesis that the rig was unbiased at the 95% confidence level. Adam jumped to conclusions too quickly.

2. Messing with Chernoff

One night, word spreads in the 2nd floor labs that there are 4 dozen free donuts upstairs. The 25 people in 271 are working on a looming midnight 61C deadline, and are each 30% likely to stay put and ignore the donuts. The 40 people in the other labs aren't as stressed, but are each 10% likely to have already filled up on pizza (per midterm 2). Assume that everyone makes their decision independently.

(a) A Chernoff bound for binomial variables that one can derive from the lecture states that for the sum of independent indicator variables with expectation μ and $\alpha \geq 1$ that $Pr[X \geq \alpha\mu] \leq e^{\alpha\mu-\mu-\alpha\mu\ln\alpha}$. Use this fact to bound the probability that there'll be enough donuts for everyone (assuming, unrealistically, that no one takes seconds).

Answer: Let S be the number of students who stay behind, which is the sum of $C_1,\ldots,C_{25}$, indicators for each of the 25 61C students staying behind, plus the sum of $O_1,\ldots,O_{40}$, indicators for each of the other 40 staying behind. $\mathbf{E}[S] = \sum_{i=1}^{25} E[C_i] + \sum_{i=1}^{40} E[O_i] = 25 \cdot 0.3 + 40 \cdot 0.1 = 11.5$. The event of there being enough donuts left is equivalent to at most 48 students coming up to get one, i.e. at least 17 staying behind. The Chernoff bound for $\mu = 11.5$ and $\alpha = 17/\mu = 1.48$ yields $\Pr[S \geq 17] \leq e^{\alpha\mu-\mu-\alpha\mu\ln\alpha} = e^{17-11.5-17\ln 1.48} < 0.318311$.

Note that we *cannot* get a *lower* bound on the probability of there being enough donuts by applying Chernoff to $D = 65 - S$, the number of students who *do* go for the donuts, to upper-bound $\Pr[D > 48]$, since $48 < E[D] = 65 - 11.5 = 53.5$, so the α needed for the Chernoff bound would be less than 1, making the Chernoff bound invalid. In other words, the Chernoff bound cannot upper-bound the probability of *not* being in a tail.

(b) Part a uses a form of Chernoff bound derived by applying the Markov bound to $\alpha^{X_1+\ldots+X_n}$. What happens in this problem if you use the same procedure but start by applying the Markov bound to $2^{X_1+\ldots+X_n}$? Hint: You should use the inequality $e^x \geq x+1$.

Answer: Following the approach from class, we observe that $f(x) = 2^x$ is monotonic and non-negative, so $x > y$ if and only if $2^x > 2^y$. Thus by the Markov bound,

$$\Pr[\sum_i X_i \geq \alpha\mu] = \Pr[2^{\sum_i X_i} \geq 2^{\alpha\mu}] \leq \frac{\mathbf{E}[2^{\sum_i X_i}]}{2^{\alpha\mu}}$$

Since the X_i's are independent:

$$\leq \frac{\prod_i \mathbf{E}[2^{X_i}]}{2^{\alpha\mu}} = \frac{\prod_i 2 \cdot p_i + 1 \cdot (1-p_i)}{2^{\alpha\mu}} = \frac{\prod_i p_i + 1}{2^{\alpha\mu}}$$

Since $e^x \geq 1+x$:

$$\leq \frac{\prod_i e^{p_i}}{e^{\alpha\mu\ln 2}} = \frac{e^{\sum_i p_i}}{e^{\alpha\mu\ln 2}} = \frac{e^{\mu}}{e^{\alpha\mu\ln 2}} = e^{\mu-\alpha\mu\ln 2}$$

In this case, this yields $\Pr[S \geq 17] \leq e^{11.5-17\ln 2} \leq 0.7532$, a substantially weaker bound. With some algebra and a bit of calculus, you can actually demonstrate that using α as the base always yields the strongest bound. This is left as an exercise to the reader.

3. It Catches Up With You

Let $X_1, \ldots, X_n$ be independent Bernoulli random variables that each take value 1 with probability p and 0 with probability $1-p$. You have learned how to use Chebyshev's inequality to say things about the probability that the sum $S = X_1 + X_2 + \cdots + X_n$ deviates from its mean (pn). In this question you will derive another bound called Chernoff's inequality that is much stronger in most cases.

(a) As an example to help you understand the setting better, assume that X_i is the outcome of a coin flip (that is $X_i = 1$ if the coin flip results in heads and otherwise $X_i = 0$). Then $p = 1/2$ and S is the number of heads you observe. Assume that $n = 100$ is the number of coin flips. The expected number of heads you see is $pn = 50$. The exact probability that $S \geq 80$ is $5.5795 \cdot 10^{-10}$. Now using Chebyshev's inequality find an upper bound for this probability. Is your upper bound much larger than the value you computed?

Answer: From Chebyshev's inequality:

$$Pr(S \geq 80) = Pr(S - 50 \geq 30) \leq Pr(|S-50| \geq 30) \leq \frac{100 \cdot 0.5 \cdot 0.5}{30^2} = 0.0278$$

This is a very conservative upper bound.

(b) Back to the general setting, prove that if $f : \{0,1\} \to \mathbb{R}$ is any function, then $f(X_1), \ldots, f(X_n)$ are independent. Hint: write down the definition of independence. If f takes the same value at 0 and 1 then everything should be obvious. It remains to prove it in the case where $f(0) \neq f(1)$.

Answer: For $f(X_1) \ldots f(X_n)$ to be independent, we need

$$Pr[f(X_1) = a_1, f(X_2) = a_2, \ldots, f(X_n) = a_n] = Pr[f(X_1) = a_1] \cdot Pr[f(X_2) = a_2] \ldots Pr[f(X_n) = a_n]$$

If f is a constant function, i.e. $f(0) = f(1) = c$, then this is trivially true as the probability is either 1 or 0 depending on the values of a_i.

So, let's look at when it's not a constant function. Say we have $f(0) = c_0$ and $f(1) = c_1$. Further, since there are the only two values our function can take, we will restrict the codomain of our function to be $\{c_0, c_1\}$. f is thus a bijective function from $\{0,1\}$ to $\{c_0, c_1\}$.

This allows us to define an inverse of f as follows: $f^{-1}(c_0) = 0$ and $f^{-1}(c_1) = 1$. With this, we can say $Pr[f(X_i) = a] = Pr[X_i = f^{-1}(a)], a \in \{c_0, c_1\}$

Now, we have:

$$\begin{aligned} Pr[f(X_1) = a_1, f(X_2) = a_2, \ldots, f(X_n) = a_n] &= Pr[X_1 = f^{-1}(a_1), X_2 = f^{-1}(a_2), \ldots, X_n = f^{-1}(a_n)] \\ &= Pr[X_1 = f^{-1}(a_1)] \cdot Pr[X_2 = f^{-1}(a_2)] \ldots Pr[X_n = f^{-1}(a_n)] \\ &= Pr[f(X_1) = a_1] \cdot Pr[f(X_2) = a_2] \ldots Pr[f(X_n) = a_n] \end{aligned}$$

Thus, we have that the $f(X_i)$s are independent.

(c) Now if we fix a number t and let $f(x) = e^{tx}$, then $f(X_i) = e^{tX_i}$. Compute the expected value of $f(X_i) = e^{tX_i}$ and write it in terms of p and t.

Answer:

$E[f(X_i)] = = pe^{t\cdot 1} + (1-p)e^{t\cdot 0}$
$= pe^t + 1 - p$
$= p(e^t - 1) + 1$

(d) The following is a famous inequality about real numbers: $1+x \le e^x$. Another variant of the inequality (which can be derived by replacing x by $x-1$) is the following: $x \le e^{x-1}$. Apply the latter inequality with x being the expected value you computed in the previous step in order to get an upper bound on $E[f(X_i)]$. (You don't need to prove either of these inequalities.)

Answer: $E[f(X_i)] = p(e^t - 1) + 1$

$\le e^{(p(e^t-1)+1)-1}$

$= e^{p(e^t-1)}$

(e) Remembering that $f(X_1), \dots, f(X_n)$ are all independent what is $E[f(X_1)f(X_2)\dots f(X_n)]$ in terms of $E[f(X_1)], \dots, E[f(X_n)]$? Use the upper bound you got from the previous step to get an upper bound on $E[f(X_1)f(X_2)\dots f(X_n)]$. You should be able to express your answer in terms of p, n, and t. Now let $\mu = pn$ be the expected value of S. Re-express your upper bound in terms of μ and t (i.e. remove the occurrences of p and n and rewrite them in terms of μ).

Answer:

$E[f(X_1)f(X_2)\dots f(X_n)] = E[f(X_1)] \cdot \dots \cdot E[f(X_n)]$
$\le e^{p(e^t-1)} \cdot \dots \cdot e^{p(e^t-1)}$

$= (e^{p(e^t-1)})^n$

$= e^{np(e^t-1)}$

$= e^{\mu(e^t-1)}$

(f) Observe that $f(X_1)\dots f(X_n) = e^{t(X_1+\cdots+X_n)} = e^{tS}$. Let us call e^{tS} the random variable Y. Does it always take positive values? Let's say we are interested in bounding the probability that $S \ge (1+\alpha)\mu$ where α is a non-negative number. Prove that $S \ge (1+\alpha)\mu$ is the same event as $Y \ge e^{t\mu(1+\alpha)}$. Use Markov's inequality on the latter event to derive an upper bound for $\Pr[S \ge (1+\alpha)\mu]$ in terms of μ, t, and α.

Answer: The exponential function always takes positive values as long as the exponent is a real number.

We know that e^{tS} is a monotonically increasing function of S, if t is positive. This means that $a \ge b$ if and only if $e^{ta} \ge e^{tb}$.

Thus, the event $S \ge (1+\alpha)\mu$ is the same as $e^{tS} \ge e^{t(1+\alpha)\mu}$ which is just $Y \ge e^{t(1+\alpha)\mu}$.

Using Markov's inequality, we have:

$Pr[S \ge (1+\alpha)\mu] = Pr[Y \ge e^{t(1+\alpha)\mu}]$

$\le \frac{E[Y]}{e^{t(1+\alpha)\mu}}$

$= \frac{E[e^{tS}]}{e^{t(1+\alpha)\mu}}$

$$= \frac{E[e^{t(X_1+\ldots+X_n)}]}{e^{t(1+\alpha)\mu}}$$

$$\leq \frac{e^{\mu(e^t-1)}}{e^{t(1+\alpha)\mu}}$$

$$= e^{\mu(e^t-1)-\mu t(1+\alpha)}$$

(g) For different values of t you get different upper bounds for the probability that $S \geq (1+\alpha)\mu$. But of course all of them are giving you an upper bound on the same quantity. Therefore it is wiser to pick a t that minimizes the upper bound. This way you get the tightest upper bound you can using this method. Assuming that α is fixed, find the value t that minimizes your upper bound. For this value of t what is the actual upper bound? Your answer should only depend on α and μ. Hint: in order to minimize a positive expression you can instead minimize its ln. Then you can use familiar methods from calculus in order to minimize the expression.

Answer: We want to minimize $e^{\mu(e^t-1)-\mu t(1+\alpha)}$ with respect to t.
This is the same as minimizing $\log(e^{\mu(e^t-1)-\mu t(1+\alpha)}) = \mu(e^t-1) - \mu t(1+\alpha)$ with respect to t.
So, let's take $g(t) = \mu(e^t-1) - \mu t(1+\alpha)$. We want $g'(t) = 0$.

$$g'(t) = \mu e^t - \mu(1+\alpha) = 0$$

$$\therefore e^t = (1+\alpha) \Rightarrow t = \log(1+\alpha)$$

We see that $g''(t) = \mu e^t > 0$, so this is indeed a minimum.

The upper bound we get with this t is:

$$e^{\mu((1+\alpha)-1)-\mu\log(1+\alpha)(1+\alpha)} = \frac{e^{\mu\alpha}}{(e^{\log(1+\alpha)})^{\mu(1+\alpha)}}$$

$$= \frac{e^{\mu\alpha}}{(1+\alpha)^{\mu(1+\alpha)}}$$

(h) Here we want to compare Chernoff's bound and the bound you can get from Chebyshev's inequality. Assume for simplicity that $p = 1/2$, so $\mu = n/2$.

First compute Chernoff's bound for the probability of seeing at least 80 heads in 100 coin flips (the quantity you bounded in the first part). Compare your answer to that part and see which one is closer to the actual value.

Now back to the setting with general n and α, write down the Chernoff bound as c^n where c is an expression that only contains α and not n. This shows that for a fixed value of α, Chernoff's bound decays exponentially in n. Now write down Chebyshev's inequality to bound $\Pr[|S-\mu| \geq \alpha\mu]$. Show that this is also a bound on $\Pr[S \geq (1+\alpha)\mu]$. Write down this bound as γn^β where γ and β are some numbers that do not depend on n. This shows that Chebyshev's inequality decays like n^β. In general an exponential decay (which you get from Chernoff's) is much faster than a polynomial decay (the one you get from Chebyshev's).

Answer: We want to find $P(S \geq 80) = P(S \geq (1+0.6)50)$. So, we have $\alpha = 0.6$ and $\mu = 50$.

Plugging in these values, we get a bound of $5.0031 \cdot 10^{-4}$. This is a much better bound than what we got from Chebyshev's inequality.

Now, let's try to get our Chernoff bound in the form c^n.

$$\frac{e^{\mu\alpha}}{(1+\alpha)^{\mu(1+\alpha)}} = \frac{e^{np\alpha}}{(1+\alpha)^{np(1+\alpha)}} = \left(\frac{e^{p\alpha}}{(1+\alpha)^{p(1+\alpha)}}\right)^n$$

So, we have $c = \dfrac{e^{p\alpha}}{(1+\alpha)^{p(1+\alpha)}}$.

Note: Using the derivation of the Chernoff bound without the approximation in part 4 (as given in note 19b), we get a much more accurate bound of $4.258 \cdot 10^{-9}$.

Now, let's look at Chebyshev's bound. First we have:

$$\begin{aligned}
P(|S-\mu| \geq \alpha\mu) &= P(S-\mu \geq \alpha\mu) + P(\mu - S \geq \alpha\mu) \\
&\geq P(S-\mu \geq \alpha\mu) \\
&= P(S \geq (1+\alpha)\mu)
\end{aligned}$$

So, from Chebyshev's inequality, we have:

$$P(S \geq (1+\alpha)\mu) \leq P(|S-\mu| \geq \alpha\mu) \leq \frac{np(1-p)}{(\alpha np)^2} = \frac{1-p}{p\alpha^2} \cdot n^{-1}$$

This gives us $\gamma = \dfrac{(1-p)}{p\alpha^2}$ and $\beta = -1$.

4. Disguise it

Collecting statistics about sensitive issues (such as the percentage of the population that have a certain STD, etc.) is always a challenge.

Such a situation can arise in 70 if the professor wants to ask people if the homeworks have been too hard recently. People who respond to such questions might be more comfortable answering the truth if the polling mechanism gives them plausible deniability. Suppose that you want to ask a Yes/No question. You ask people to first roll a dice (on their own). If the result is 6 they should report the true answer, but otherwise you ask them to flip a coin and based on that randomly answer Yes/No. The dice roll is kept secret and not revealed to the professor.

(a) First, let's consider the system without the dice-rolling part nor the coin-tossing part. Suppose that exactly $\frac{5}{6}$ of the students are told to give a canned answer and exactly half of the canned answers are Yes and half of the canned answers are No. The remaining $\frac{1}{6}$ of the students sampled give the true answer. Further assume that these $\frac{1}{6}$ students have exactly the same proportion of YES/NO as the whole student population.

The professor doesn't know which students were canned and which were giving real answers.

Suppose that a fraction q of the answers you get are Yes. What fraction p of the population should you assume would answer the original question with a Yes (assuming sensitivity was not an issue)? Express this as a formula in terms of q.

Answer: A $5/6$ fraction of the students give canned answers. Out of this $5/6$ fraction exactly half answer yes. So a fraction of $\frac{5}{6} \times \frac{1}{2} = \frac{5}{12}$ out of all students give canned yes answers.

A fraction $1/6$ of the students give the correct answer. So exactly a p fraction of them answer Yes. Therefore a fraction $p/6$ of all students give true yes answers.

Every yes answer the professor receives is either a canned yes answer or a true yes answer. $5/12$ fraction of the students give the first kind and $p/6$ give the second kind. Therefore $5/12 + p/6$ fraction of the students give a yes answer. This means that

$$q = \frac{5}{12} + \frac{p}{6}$$

Rearranging this equality gives us

$$p = 6 \times (q - \frac{5}{12}) = 6q - \frac{5}{2}$$

So we should assume $6q - 2.5$ fraction of the population would have answered yes.

(b) Now, suppose only that coin toss has been introduced back into the problem, but everything else is as before. Exactly $\frac{5}{6}$ of the students toss a coin for their answers while the remaining $\frac{1}{6}$ of students answer honestly. Furthermore, this subset of students has exactly the same proportion of YES/NO as the entire population.

Argue using the law of large numbers that as the number of people asked goes to infinity the formula from the previous part approaches the true fraction with confidence approaching 1.

Use your calculations/simulations to say how big of a class must it be so that we believe that we will get p correct to within ± 0.1 with a confidence of 95%?

Answer: This situation is similar to that of bias estimation for coin tosses. As the number of students approaches infinity, the number of coin tossing students polled also approaches infinity. Each one of them answers with a coin toss, therefore the fraction of the randomized answers that respond with a Yes approaches $\frac{1}{2}$.

Now if out of the randomized answers, a fraction r of them answer yes, then this fraction counted out of the total number of students polled would be $\frac{5r}{6}$. The number of non-randomized yes answers would still be $\frac{p}{6}$. So q would be $\frac{5r+p}{6}$.

Last part tells us that our estimate of p should be $6q - 5/2$. Plugging in $\frac{5r+p}{6}$ for q tells us that the estimate is $6 \times \frac{5r+p}{6} - \frac{5}{2} = p + (5r - \frac{5}{2})$.

One can see that with confidence approaching one, the value r approaches $\frac{1}{2}$ (i.e. gets arbitrarily close to it). Therefore the value $5r - \frac{5}{2}$ approaches $5 \times \frac{1}{2} - \frac{5}{2} = 0$ with confidence approaching one.

The difference between our estimate of p and the real value of p is always $5r - \frac{5}{2}$. For this number to be at most 0.1 in absolute value, we must have

$$|r - 0.5| \leq 0.02 \iff r \in [0.48, 0.52]$$

Now let us go back to what we know about Galton-Watson processes. We know that if you flip a coin m times and count $+1$ for heads and -1 for tails then roughly 95% of the time you would get a number in the range $[-2\sqrt{m}, 2\sqrt{m}]$. Here $m = 5n/6$ is the number of people with randomized answers.

How does this sum of ± 1's relate to r? Well if r fraction of the answers are yes, then r fraction of the numbers are $+1$ and $1-r$ fraction are -1. Therefore the sum of ± 1's is $rm-(1-r)m=(2r-1)m$.

$$|(2r-1)m| \leq 2\sqrt{m} \iff |r-0.5| \leq \frac{1}{\sqrt{m}}$$

So if we make sure that $\frac{1}{\sqrt{m}} \leq 0.02$, then $|r-0.5| \leq 0.02$ with 95% confidence which is exactly what we wanted.

For $\frac{1}{\sqrt{m}} \leq 0.02$ to happen we must have $m \geq 2500$. This translates to $n \geq 6m/5 = 3000$.

(c) Now consider the original scheme with the dice and the coin. Argue using the law of large numbers that as the number of people asked goes to infinity the formula from the previous part approaches the true fraction with confidence approaching 1.

(We are not asking for proofs here because the Laws of Large Numbers are Empirical Facts for now. However, you should try to be precise in your argumentation. Later on in the course, we will be able to prove such statements.)

Answer: For each person asked the probability of receiving a yes can be calculated as follows: with $5/6$ probability that person is going to give a canned answer which with probability $1/2$ is going to turn out yes. So with probability $\frac{5}{6} \times \frac{1}{2}$ the answer is a canned yes. With probability $\frac{1}{6}$ however the answer is honest and then because this is a randomly sampled person, the proabability of receiving a yes would be p. So with a probability of $p/6$ we receive an honest yes. So overall the probability of receiving a yes from a randomly picked person is $\frac{5}{12} + \frac{p}{6}$.

Now using the law of large numbers one sees that as the number of persons polled approaches infinity, the fraction of them that answer yes must approach $\frac{5}{12} + \frac{p}{6}$ with confidence 1. If we name the value $\frac{5}{12} + \frac{p}{6}$ as $\hat{q}$, then this means that $|\hat{q} - q|$ gets arbitrarily small with confidence approaching one as the number of persons asked approaches infinity. But multiplication by 6 and subtraction of 2.5 are both continuous operations. So $6q - 2.5$ gets arbitrarily close to $6\hat{q} - 2.5 = p$ with confidence approaching one.

5. Probabilistically Buying Probability Books

Chuck will go shopping for probability books for K hours. Here, K is a random variable and is equally likely to be 1, 2, or 3. The number of books N that he buys is random and depends on how long he shops. We are told that

$$\Pr[N=n|K=k] = \frac{c}{k}, \qquad \text{for } n = 1, \ldots, k$$

for some constant c.

(a) Compute c.

Answer: For any k, we know that probabilities conditioned on $K=k$ must sum to 1, i.e

$$\sum_{n} \Pr[N=n|K=k] = 1 \quad ,$$

so it must be that

$$1 = \sum_{n=1}^{k} \Pr[N=n|K=k] = k \times \frac{c}{k} = c \quad .$$

Thus, $c=1$.

(b) Find the joint distribution of K and N.

Answer: The joint distribution specifies $\Pr[N=n\cap K=k]$ for all n and k. Note that

$$\Pr[N=n\cap K=k]=\Pr[N=n|K=k]\Pr[K=k]$$

and we know $\Pr[N=n|K=k]$ and $\Pr[K=k]$ (it says all $k\in\{1,2,3\}$ are equally likely). We use this formula to calculate $\Pr[N=n\cap K=k]$ for each n,k and list the result in a table:

$n\setminus k$	1	2	3
1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{9}$
2	0	$\frac{1}{6}$	$\frac{1}{9}$
3	0	0	$\frac{1}{9}$

(c) Find the marginal distribution of N.

Answer: The marginal distribution of N is given by the value of $\Pr[N=n]$, for each possible value of n. By the total probability rule,

$$\Pr[N=n]=\Pr[N=n\cap K=1]+\Pr[N=n\cap K=2]+\Pr[N=n\cap K=3]\ \ .$$

Thus, we get

$$\Pr[N=n]=\begin{cases}\frac{1}{3}+\frac{1}{6}+\frac{1}{9} & \text{if } n=1\\ \frac{1}{6}+\frac{1}{9} & \text{if } n=2\\ \frac{1}{9} & \text{if } n=3\end{cases}=\begin{cases}\frac{11}{18} & \text{if } n=1\\ \frac{5}{18} & \text{if } n=2\\ \frac{2}{18} & \text{if } n=3\end{cases}$$

(d) Find the conditional distribution of K given that $N=1$.

Answer: By definition, $\Pr[K=k|N=1]=\frac{\Pr[K=k\cap N=1]}{\Pr[N=1]}$. The numerator comes from the joint distribution of N and K (part (b)), and the denominator comes from the marginal distribution of N (part (c)). Plugging in, we get

$$\Pr[K=k|N=1]=\begin{cases}\frac{\frac{1}{3}}{\frac{11}{18}} & \text{if } k=1\\ \frac{\frac{1}{6}}{\frac{11}{18}} & \text{if } k=2\\ \frac{\frac{1}{9}}{\frac{11}{18}} & \text{if } k=3\end{cases}=\begin{cases}\frac{6}{11} & \text{if } k=1\\ \frac{3}{11} & \text{if } k=2\\ \frac{2}{11} & \text{if } k=3\end{cases}$$

(e) We are now told that he bought at least 1 but no more than 2 books. Find the conditional mean and variance of K, given this piece of information.

Answer: We first compute the distribution $\Pr[K=k|N=1\cup N=2]$ as we did in part (d):

$$\Pr[K=k|N=1\cup N=2]=\begin{cases}\frac{\frac{1}{3}}{\frac{16}{18}} & \text{if } k=1\\ \frac{\frac{1}{6}+\frac{1}{6}}{\frac{16}{18}} & \text{if } k=2\\ \frac{\frac{1}{9}+\frac{1}{9}}{\frac{16}{18}} & \text{if } k=3\end{cases}=\begin{cases}\frac{3}{8} & \text{if } k=1\\ \frac{3}{8} & \text{if } k=2\\ \frac{2}{8} & \text{if } k=3\end{cases}$$

Now, the mean will be

$$\mathbb{E}(K|N=1\cup N=2)=1\times\frac{3}{8}+2\times\frac{3}{8}+3\times\frac{2}{8}=\frac{15}{8}$$

and the variance will be

$$\begin{aligned}
\operatorname{Var}(K|N=1\cup N=2) &= \mathbb{E}(K^2|N=1\cup N=2)-\mathbb{E}(K|N=1\cup N=2)^2\\
&= 1^2\times\frac{3}{8}+2^2\times\frac{3}{8}+3^2\times\frac{2}{8}-\left(\frac{15}{8}\right)^2\\
&= \frac{39}{64}\\
&\approx 0.61
\end{aligned}$$

(f) The cost of each book is a random variable with mean 3. What is the expectation of his total expenditure? *Hint:* Condition on events $N=1,\dots,N=3$ and use the total expectation theorem.

Answer: Let X be his total expenditure. Using the total expectation theorem, we have

$$\mathbb{E}(X)=\mathbb{E}(X|N=1)\Pr[N=1]+\mathbb{E}(X|N=2)\Pr[N=2]+\mathbb{E}(X|N=3)\Pr[N=3]$$

Since each book has an expected price of 3, $\mathbb{E}(X|N=n)=3\times n$, giving

$$\begin{aligned}
\mathbb{E}(X)=\mathbb{E}(X|N=1)\Pr[N=1]+\mathbb{E}(X|N=2)\Pr[N=2]+\mathbb{E}(X|N=3)\Pr[N=3]\\
=3\times\frac{11}{18}+6\times\frac{5}{18}+9\times\frac{2}{18}\\
=\frac{9}{2}\ .
\end{aligned}$$