Math 394

Solutions to Homework Assignment

Problems from Chapter 5

Part I
Assignment Problems

#1 (C)

The point of this problem is that a (nice) function f can be considered a probability density if (and only if)

1. $f(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f(x)dx = 1$ (if f is defined over, say, and interval $[a, b]$, it just means that it is taken to be identically equal to 0 for x outside that interval)

In this case, we have

1. $x^2 \geq 1$ when $-1 \leq x \leq 1$
2. $$\int_{-\infty}^{\infty} f(x)dx = \int_{-1}^{1} c\left(1 - x^2\right) dx = \left[c\left(x - \frac{x^3}{3}\right)\right]_{-1}^{1} =$$
$$= c\left(1 - (-1) - \frac{1}{3} + \left(-\frac{1}{3}\right)\right) = c\left(2 - \frac{2}{3}\right) = c \cdot \frac{4}{3}$$
This will be equal to 1 if $c = \frac{3}{4}$.

The cumulative distribution function is now easy to calculate. Since $-1 < X < 1$,

- $F_X(x) = 0$ for $x \leq -1$
- $F_X(x) = 1$ for $x \geq 1$
- For $-1 < x < 1$,

$$F_X(x) = P[X \leq x] = \int_{-\infty}^{x} f(t)\,dt = \int_{-1}^{x} \frac{3}{4}\left(1 - t^2\right) dt =$$

$$\frac{3}{4}\left[t - \frac{t^3}{3}\right]_{-1}^{x} = \frac{3}{4}\left(x - \frac{x^3}{3} + 1 - \frac{1}{3}\right) = \frac{1}{2} - \frac{x^3}{4} + \frac{3}{4}x$$

Looking back at the determination of c, with a question like this, you will be given a density in the form of an unknown constant multiplying an explicit function, as in

$$f(x) = c \cdot g(x)$$

where $g(x) \geq 0$. To determine c, you will then evaluate the *number* $\int_{-\infty}^{+\infty} g(x)\,dx = K$ (which must be finite for the problem to have a solution). For f to be a density, we will then need $c = \frac{1}{K}$

Of course, for this method to work, $g(x)$ must be nonnegative, and integrable over the whole line! This is no different from the similar situation with discrete, but not finite, RVs: a given sequence of numbers $a_i : i = 1, 2, \ldots$ can be used to define a probability distribution, if $a_i \geq 0$, and $\sum_{i=1}^{\infty} a_i = a < \infty$: the probability distribution is then given by the sequence $p_i = \frac{a_i}{a}$.

#3

The scheme is exactly the same as in problem 1: we need to make sure that the function is nonnegative, and integrable. Now

$$2x - x^3 \geq 0$$

if $x \geq 0$ and $x^2 \leq 2$, i.e., in $[0, \sqrt{2}]$. Since $\frac{5}{2} > \sqrt{2}$ this is not a candidate for probability density.

Similarly,

$$2x - x^2 > 0$$

if $0 \leq x$, and $2 \geq x$. But $\frac{5}{2} = 2.5 > 2$ and this function is also not a suitable candidate.

#4 (C)

We apply the definitions:

a Thus,

$$P[X > 20] = \int_{20}^{\infty} f(x)dx = \int_{20}^{\infty} \frac{10}{x^2}dx = \lim_{K\to\infty} \left[-\frac{10}{x}\right]_{20}^{K} = 0 + \frac{1}{2} = 0.5$$

b The cdf is given by $F(x) = \int_{-\infty}^{x} f(t)dt$, hence

$$x < 10 : F(x) = 0$$

$$x \geq 10 : F(x) = \int_{10}^{x} \frac{10}{t^2}dt = \left[-\frac{10}{t}\right]_{10}^{x} = 1 - \frac{10}{x}$$

(note that F is continuous at $x = 10$ - hence it makes no difference whether we define $F = 0$ for $x < 10$ or $x \geq 10$, and similarly for the definition of F "from the right". If F had had a discontinuity, convention is to define it as *continuous from the right.*

c Now we have 6 RVs, and we haven't been told anything about their *joint* distribution. We need to assume that they behave independently (which is by no means an obvious assumption, but we are given no choice), so that we can treat this as a 3-out-of-6 system, with probability of failure given, for each component independently, by

$$P[X \leq 15] = 1 - \frac{10}{15} = \frac{1}{3} = \bar{p}$$

(note that for continuous RVs, it makes no difference if we include or exclude the endpoints of intervals, since the probability of X being exactly equal to an endpoint is zero - or, equivalently, because, for integrable functions

$$\lim_{x \to b} \int_a^x f(t)dt = \int_a^b f(t)dt$$

i.e., antiderivatives are continuous). Applying the argument we met in the previous assignment, the requested probability is given by

$$\sum_{k=3}^{6} \left(\begin{array}{c} 6 \\ k \end{array} \right) (1 - \bar{p})^k \, \bar{p}^{6-k} =$$

$$= 20 \cdot \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^3 + 15 \cdot \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^2 + 6 \cdot \left(\frac{2}{3}\right)^5 \left(\frac{1}{3}\right) + \left(\frac{2}{3}\right)^6 =$$

$$= \frac{2^3}{3^6} \left(20 + 15 \cdot 2 + 6 \cdot 2^2 + 2^3\right) = \frac{8}{729} (20 + 30 + 24 + 8) = \frac{8 \cdot 82}{729} = .89986$$

#6

Here, we use the formula

$$EX = \int_{-\infty}^{\infty} x f_X(x) dx$$

which is meaningful only if $x f_X$ is an integrable function (for instance, a density of the form $\frac{c}{x^2}$ for $x \geq 1$, with suitable c, would be indeed a density, but it would have no expected value. More to the point, the Cauchy distribution (with density of the form $\frac{c}{a^2+x^2}$), which appears in several applications, also has no expected value).

a First, note that f is indeed a density function (it is a member of the family of so-called "Gamma" distributions) :

$$\int_0^{\infty} \frac{1}{4} x e^{-\frac{x}{2}} dx = \frac{1}{4} \left(\left[-2xe^{-\frac{x}{2}} \right]_0^{\infty} + 4 \right) = 1$$

To compute EX, we calculate

$$\int_{-\infty}^{\infty} x f(x) dx = \frac{1}{4} \int_0^{\infty} x^2 e^{-\frac{x}{2}} dx$$

by integrating by parts twice. To streamline the calculations, it doesn't hurt to remember (from the exponential distribution - cf. p.209 in the book) that

$$\int_0^\infty ke^{-kx}dx = 1, \quad \int_0^\infty kxe^{-kx}dx = \frac{1}{k}, \quad \int_0^\infty kx^2e^{-kx}dx = \frac{2}{k^2}$$

It follows that $EX = 4$

b We already saw in problem 1 that $c = \frac{3}{4}$. As for the expected value,

$$EX = \frac{3}{4}\int_{-1}^{1} x\left(1-x^2\right)dx = \frac{3}{4}\left[\frac{x^2}{2} - \frac{x^2}{4}\right]_{-1}^{1} = 0$$

Of course, this could have been seen immediately, since we are integrating an odd function ($f(x) = -f(-x)$) over an interval symmetric around 0.

c First of all, this function is a density, since

$$\int_5^\infty \frac{5}{x^2} = \left[-\frac{5}{x}\right]_5^\infty = 1$$

but it has no expected value: $x \cdot \frac{5}{x^2} = \frac{5}{x}$ is notoriously not integrable at $+\infty$ ($\int_1^X \frac{5}{x}dx = \log X \to \infty$ as $X \to \infty$).

#7

A we noted in problem 1, a density function has to satisfy $f \geq 0$, and $\int_{-\infty}^{\infty} f dx = 1$. The last formula provides one relation to determine the missing constants. To determine both we need another relation, and that's provided by the value of $\int_{-\infty}^{\infty} xf dx$. We have two equations in two unknowns a and b:

$$\int_0^1 \left(a + bx^2\right)dx = a + \frac{b}{3} = 1$$

$$\int_0^1 x\left(a + bx^2\right)dx = \frac{a}{2} + \frac{b}{4} = \frac{3}{5}$$

Solving this simple system gives $a = \frac{3}{5}, b = \frac{6}{5}$

#9

Let's go to example 4b on p.130 of the book. Recapping, we consider a random variable X representing the number of units of a product *sold* by a department store that has to stock them in advance[1]. Suppose a stock of s has been ordered. Then a profit of b per unit sold, and a loss of l per unit left unsold

[1] The book, mysteriously, talks about X as the amount of units "ordered", rather than sold. Maybe they were thinking of a mail-order only store, but it sure makes for a confusing discussion. In any case, equation (1) (which is also the equation at the bottom of p. 134) makes clear what this is all about.

is recorded. Incidentally, we are not considering "virtual losses", due to missed sales because of insufficient stocked items that cause some demand not to be satisfied. The equation is then (cf. the first equation on page 130), calling the profit, a function of s, $P(s)$,

$$P(s) = b(s \wedge X) - (s - X)^{+} l \qquad (1)$$

(we introduced a couple of common notations: $a \wedge b = \min(a, b)$, and

$$a^{+} = \begin{cases} a & a > 0 \\ 0 & a \leq 0 \end{cases} = \max(a, 0)$$

to shorten the writing).

Now, the expected profit is

$$EP(s) = bE[s \wedge X] - lE(s - X)^{+} \qquad (2)$$

and some care is needed, since neither function of X is linear!

As you may guess by looking at the discrete variable solution on p. 135, the trick is to break up the expectation in two parts, depending on whether $s < X$ or not. Equivalently, we could introduce two RVs

$$Y = s \wedge X; Z = (s - X)^{+}$$

and write down their densities (actually, they are "mixed" random variable: they take a specific value with a positive probability, while the remaining values are continuous:

$$P[Y = s] = P[X \geq s]; P[Z = 0] = P[X \geq s]$$

and a density elsewhere) while, conditioning on $X \leq s$, we have a (conditional) density, the density of X, suitably normalized.

There are tricks to represent even the distribution of discrete RVs in terms of "densities" (recall that some books refer to the pmf of a discrete RV as a "discrete density"). This is done in a handwaving manner by introducing the so called "$\delta-$function" (which is not a function at all) $\delta(x - a)$, where δ is characterized by the property that, for any continuous function,

$$\int_{-\infty}^{\infty} f(x)\,\delta(x - a)\,dx = f(a)$$

which is sometimes "explained" in physics and engineering books by stating that $\delta$is "equal to zero everywhere, except at a, where it is infinite" (there are rigorous mathematical ways to define such an object - as a Schwartz "distribution" - no relation with probability distributions, of course - or as a measure, or as a linear functional on an appropriate function space - all the preceding terms are, in a sense, equivalent or closely connected). Hence, a pmf $p_1, p_2, \ldots, p_n$, for values $x_1, x_2, \ldots, x_n$ can be written (formally) as

$$f(x) = \sum_{k=1}^{n} p_k \delta(x - x_k) \qquad (3)$$

This is only meant to show that the following solution for the problem, assuming X is a continuous RV, applies to the discrete (or even mixed) case as is, using the formal trick of equation (3).

We compute the expected values in (2):

$$E[s \wedge X] = sP[X > s] + \int_0^s xf(x)\,dx$$

(here f is the density of X)

$$E\left[(s - X)^+\right] = \int_0^s (s - x) f(x)\,dx$$

so that the expected profit is

$$bsP[X > s] + b\int_0^s xf(x)\,dx - lsP[X \le s] + l\int_0^s xf(x)\,dx$$

Let's derive this expression with respect to s (noting that $P[X > s] = 1 - F(s)$, and $[X \le s] = F(s)$, where F is the cdf of X, and that $F' = f$):

$$b(1 - F(s)) - bsf(s) + bsf(s) - lF(s) - lsf(s) + lsf(s) =$$

$$= b - (b + l)F(s)$$

which is zero for any s^* such that

$$F(s^*) = \frac{b}{b + l} \tag{4}$$

Note that there could be more than one s^*, if the function F happened not to be strictly monotone. However, for instance in the discrete case, only the left endpoint of the interval of s satisfying (4) would have physical meaning (if you are counting single units, you cannot sell "2.45 single units"). This has been expressed as an "inequality" condition in the solution to the discrete problem

#10

a We have a density function

$$f_p(t) = \begin{cases} 1 & 7 \le t \le 8 \\ 0 & \text{otherwise} \end{cases}$$

for the arrival time of the passenger. The passenger will take the train to B if he/she arrives between 7 and $7:05$, $7:15$ and $7:20$, and so on, up to

7 : 45 to 7 : 50, for a total time "window" of 20 minutes. If he/she arrives in the remaining time intervals, the "winning" train goes to A, and that is over a total "window" of 40 minutes. Hence, the probability of going to A is $\frac{40}{60} = \frac{2}{3}$. Incidentally, the probability of the passenger catching the train that leaves at 7:00 is zero, since $\int_0^0 f_p(t)dt = 0$.

b At first, we have 5 minutes of opportunity to get the next train, which now is the one going to A. After this initial period, the passenger has again 10-minute windows to catch a train to A, alternating with 5-minute windows to catch a train to B, and, if late, will catch the 8:00 train to A. Hence the percentage for the two destinations is the same as before: the probability of going to A is $\frac{2}{3}$.

#13 (C)

a Since you will wait for more than 10 minutes if the bus arrives later than 10:10, the probability is the probability of such an event. Since the arrival time distribution is uniform over $[10:00, 10:30]$, we see that the probability is $\frac{2}{3}$.

b Now we have a conditional problem: if W is the waiting time (which coincides with the arrival time counting from 10:00, and hence is uniformly distributed over $[0, 30]$) in minutes,

$$P\left[W > 25 \,|W > 15\right] = \frac{P\left[W > 25 \cap W > 15\right]}{P\left[W > 15\right]} = \frac{P\left[W > 25\right]}{P\left[W > 15\right]} = \frac{\frac{5}{30}}{\frac{15}{30}} = \frac{1}{3}$$

(which is almost obvious: you are wondering if the bus will come in the last third of your remaining time window).

Note Note that, as you push the question to waiting times closer and closer to 30, your probability, conditional or not, of waiting that long goes to zero (in fact, $P\left[W > 30\right] = 0$). If the bus arrival distribution had been exponential, things would be much different: while your probability of an early bus would in fact be higher, once you had waited 15 minutes, the arrival probability would not have increased at all! In fact, the exponential distribution, just like the geometric in the discrete case, does not improve your odds, as time goes by without anything happening (cf. the connection between the two, that we showed in class). For example, if W was exponential, with mean value 15 (that's the mean value of the W in the problem), it would have density, for $t > 0$ equal to

$$\frac{1}{15}e^{-\frac{t}{15}}$$

and your probability of not waiting more than, say, 10 minutes, would be

$$F(10) = 1 - e^{-\frac{10}{15}} = 1 - e^{-\frac{2}{3}} = .48658$$

much better than the $\frac{1}{3}$ that would happen with the uniform W. On the other hand,

$$P[W > 25 \,|W > 15] = \frac{P[W > 25 \cap W > 15]}{P[W > 15]} = \frac{P[W > 25]}{P[W > 15]} =$$

$$= \frac{e^{-\frac{25}{15}}}{e^{-\frac{15}{15}}} = e^{-\frac{10}{15}} = .48658$$

i.e., your probability of waiting 10 minutes or more has not improved one bit. This *not* the same situation as before: $P[W > 30] \neq 0$, and, in fact, by the same calculation,

$$P[W > 39 \,|W > 29] = .48658$$

as opposed to zero.

#15 (C)

The trick for these computations, which are critical in statistical applications is the following:

> If X is a normal RV, with expected value μ, and variance σ^2, the RV Z defined by
>
> $$Z = \frac{X - \mu}{\sigma}$$
>
> (note that the denominator is σ, *not* σ^2) is normal with expectation 0, and variance 1. Consequently, its cdf can be looked up in tables and/or coughed up by a calculator or computer program (e.g., any modern spreadsheet) that has it pre-programmed. Note that
>
> $$P[a < X < b] = P\left[\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} = Z < \frac{b - \mu}{\sigma}\right]$$

Let's apply this to the questions ($\mu = 10, \sigma^2 = 36, \sigma = \sqrt{36} = 6$):

a $P[X > 5] = P\left[\frac{X-10}{6} > \frac{5-10}{6}\right] = P\left[Z > -\frac{5}{6}\right] = .79767$

b $P[4 < X < 16] = P\left[\frac{4-10}{6} < Z < \frac{16-10}{6}\right] = P[-1 < Z < 1] = .68269$ (note the famous values for

$$P[-1 < Z < 1] \simeq .68, P[-2 < Z < 2] = .95450 \simeq .955, P[-3 < Z < 3] = .99730 \simeq .997$$

expressing the probability of a normal RV to be within $\sigma, 2 \cdot \sigma, 3 \cdot \sigma$, respectively, from its expected value).

c $P[X < 8] = P\left[Z < \frac{8-10}{6} = -\frac{1}{3}\right] = .36944$

d $P[X<20] = P\left[Z < \frac{20-10}{6} = \frac{5}{3}\right] = .95221$

e $P[X>16] = P\left[Z > \frac{16-10}{6} = 1\right] =$

$$= \frac{1}{2} P[|Z|>1] = \frac{1}{2}(1 - P[-1<Z<1]) = \frac{1}{2}(1-.68269) = .15866$$

Remark In the last calculation, we used the result from point b to save a trip to the tables. The idea was to rewrite the event $Z>1$ in terms of other events whose probability we had already looked up, relying on the symmetry of the normal distribution. If course, it might be just as fast to evaluate $1 - P[Z \leq 1]$ through the tables.

#19

This type of problem works by recasting the question in terms of a *standard* normal distribution, and then going to the tables. Thus,

$$P[X>c] = P\left[\frac{X-EX}{\sqrt{Var[X]}} > \frac{c-EX}{\sqrt{Var[X]}}\right] = P\left[\frac{X-EX}{\sqrt{Var[X]}} > \frac{c-12}{\sqrt{4}}\right] =$$

$$P\left[Z > \frac{c}{2} - 6\right]$$

where Z is a standard normal. Reading tables in reverse, we find out that

$$P[Z > z_{.1}] = .10$$

where $z_{.1} = 1.2816$. Hence,

$$\frac{c}{2} = 7.2816$$

$$c = 14.5632$$

#22 (C)

We are asked the probability that a RV distributed like $N\left(.9000, 9\cdot 10^{-6}\right)$ will take values in the interval $[.8950, .9050]$ and to determine σ^2, so that this probability will not exceed .01

a Let's call X the width of the slot. Then

$$P[.8950 < X < .9050] = P[-.0050 < X - .9000 < .0050] = P\left[\frac{-.0050}{.0030} < \frac{X-.9000}{.0030} < \frac{.0050}{.0030}\right] =$$

$$= P\left[-\frac{5}{3} < Z < \frac{5}{3}\right]$$

(where Z is a standard normal). From the tables,

$$P\left[-\frac{5}{3} < Z < \frac{5}{3}\right] = 2\Phi\left(\frac{5}{3}\right) - 1 = .90442$$

(note that the solution manual has .9050, since it rounded $\frac{5}{3}$ to 1.67, causing a relative error of roughly 10^{-3}- most likely irrelevant, since we are working in an approximation form the start) where Φ is the cdf of a standard normal. The defective forges will thus form a percentage of $1 - .90442 = .095581$, or, approximately, 9.6%

b We solve the second question by making a reverse lookup of the tables:

$$P\left[|Z| < \frac{z_{.995}}{2}\right] = .99$$

with $z_{.995} = 2.5758$. Hence,

$$P\left[\left|\frac{X - .9000}{\sigma}\right| < 2.7578\right] = .99$$

or

$$P\left[|X - .9000| < \sigma 2.5758\right] = .99$$

Now,

$$\sigma 2.5758 \leq .0050$$

if

$$\sigma \leq \frac{.0050}{2.5758} = .0019411$$

#23

The number 6 appears with probability $\frac{1}{6}$, thus the number of appearances over n throws is a binomial with parameters $n, \frac{1}{6}$. If X is this random number,

$$\frac{X - \frac{n}{6}}{\sqrt{\frac{n}{6} \cdot \frac{5}{6}}}$$

is approximately a standard normal. For $n = 1000$, $150 \leq X \leq 200$ if

$$\frac{150 - \frac{1000}{6}}{\sqrt{\frac{1000}{6} \cdot \frac{5}{6}}} \leq \frac{X - \frac{1000}{6}}{\sqrt{\frac{1000}{6} \cdot \frac{5}{6}}} \leq \frac{200 - \frac{1000}{6}}{\sqrt{\frac{1000}{6} \cdot \frac{5}{6}}}$$

i.e.

$$\frac{900 - 1000}{\sqrt{5000}} \leq \frac{X - \frac{1000}{6}}{\sqrt{\frac{1000}{6} \cdot \frac{5}{6}}} \leq \frac{1200 - 1000}{\sqrt{5000}}$$

$$-\frac{100}{50\sqrt{2}} \leq Z \leq \frac{200}{50\sqrt{2}}$$

where Z is (approximately) a standard normal. Since

$$P\left[-\sqrt{2} \leq Z \leq 2\sqrt{2}\right] = .91091$$

this is the requested probability[2].

On the other hand, if 6 appeared exactly 200 times, the remaining 800 throws were divided between the remaining 5 possible outcomes. Out of 800 throws, the number of 5s (conditional probability $\frac{1}{5}$) Y satisfies

$$P[Y < 150] = P\left[\frac{Y - 800 \cdot \frac{1}{5}}{\sqrt{800 \cdot \frac{1}{5} \cdot \frac{4}{5}}} < \cdot \frac{750 - 800}{2 \cdot \sqrt{800}}\right] =$$

$$= P\left[Z < -\frac{5}{4\sqrt{2}}\right] = .18838$$

Remark The solution manual suggests to calculate $P[149.5 < X < 200.5]$, instead of what we did above, i.e., $P[150 \leq X \leq 200]$. This is a homage to a practice (sometimes called "smoothing") that purports to take into account the fact that we are approximating a discrete distribution with a continuous one, by, somehow, interpolating by halves between the "really possible" values. This practice is not completely justified. In fact, if the number of trials is too small to invoke the Central Limit Theorem straight, it is unlikely that this correction will go in the right direction anyway. If the number of trials is adequate (and for a binomial distribution, with p not too close to 0 or to 1, such a number is really low), there is simply no point in doing so, since $\frac{.5}{\sqrt{n}}$ will in any case be of little relevance. Also, since we are doing an approximation anyway, it doesn't make too much sense to push the number of significant digits too far, which additionally makes the correction more a matter of whim that a sound mathematical need.

#26 (C)

This probability of error in our statistical reasoning is called an "error of type I" (i.e., we reject a statement about the coin, even though it is true). If X is the number of heads, using the normal approximation, we will compute, for $p = 0.5$,

$$P[X \geq 525] = P\left[\frac{X - 1000 \cdot .5}{\sqrt{1000 \cdot .25}} \geq \frac{525 - 500}{\sqrt{250}}\right] = P\left[Z \geq \frac{25}{\sqrt{250}}\right] = .056923$$

[2]The solution manual has the value .9258, for a discrepancy of about 1%. There are at least two reasons for this. One is discussed in the remark at the end of this problem. The other is that the calculations in the book are full of rounding errors, since no effort is made to postpone divisions and root extractions to the very end, nor of reducing the square root to the simplest possible. The same discrepancy is present in the second calculation.

(in statistical practice, the threshold would be set a little bit higher, since it is customary to work with a probability of error of type I of 5%).

If we now assume that $p = .55$, we will reach a false conclusion with probability

$$P[X < 525] = P\left[\frac{X - 1000 \cdot .55}{\sqrt{1000 \cdot .55 \cdot .45}} < \frac{525 - 550}{\sqrt{55 \cdot 4.5}}\right] =$$

$$P\left[Z < -\frac{25}{\sqrt{247.50}}\right] = .56018$$

Seen from the point of view of a bias in favor of the coin being fair, this is called the probability of an "error of type II", i.e. the probability of accepting the statement (in our case, "the coin is fair"), even though it is false.

This computations show that our test is reasonably strong to distinguish in a serious way between $p = .5$ and $p = .55$. Actually, in practice, people will be quite happy with a probability of error of type II no larger than 10%. If we wanted tighter numbers (i.e., smaller probabilities of error), we would have to throw more times, but there is a diminishing return here, since the reduction in standard deviation (=square root of the variance) that we obtain is only proportional to the square root of the number of throws - to get a reduction of order 10, we need 100 throws, etc. Note also that, whatever numbers should come out, our procedure will only be correct *with a certain probability.* A statistical test never *proves* anything: it only checks whether the data we actually observe are likely or not, given our assumptions.

#29 (C)

We have a the outcome of the sequence of a sequence of variables with values u or d, but in a multiplicative way: if the price today is s, the price tomorrow is either su or sd. The solution book suggests to proceed as follows: let X the number of times the stock increases, and $1000 - X$ the number of decreases. At the end of 1000 periods, the stock will have changed by a factor of

$$u^X d^{1000-X}$$

and we ask whether this is greater than 1.3 (a .3 increase). This happens if

$$u^X d^{1000-X} \geq 1.3$$

$$X \geq \frac{\log 1.3 - 1000 \log d}{\log \frac{u}{d}} = 469.2$$

and now the problem is finding the probability that a binomial RV with parameters $1000, .52$ is greater than 469.2. With the usual normal approximation, this becomes

$$P\left[Z \geq \frac{469.2 - 1000 \cdot .52}{\sqrt{1000 \cdot .52 \cdot .48}}\right] = P\left[Z \geq \frac{469.2 - 1000 \cdot .52}{\sqrt{1000 \cdot .52 \cdot .48}}\right] = P\left[Z \geq -\frac{50.800}{\sqrt{249.60}}\right] =$$

$$P[Z \geq -3.2154] = .99935$$

We can also apply our usual machinery, which is concentrated on sums, rather than products, by working with logarithms. This is not necessarily simpler or faster, but illustrates an interesting model for the stock market.

Let's call S the log of the price $Y = e^S$, and $a = \log u$, $-b = \log d$ $a > 0, b > 0$ (for financial reasons, we need to assume that $u > 1$, and $d < 1$ - otherwise, since if tomorrow's price was surely over or under today's price you would be able to make unlimited amounts of money by buying today and selling tomorrow, or vice-versa, by buying on credit or selling "short", i.e., without actually owning the stock)[3].

Hence, with probability p, S will change to $S+a$, and with probability $1-p$, to $S-b$. Hence, with probability $\binom{n}{k} p^k (1-p)^{n-k}$ the log of the price of the stock, after n days, would be $S_n = S + ka - (n-k)b$. The expected value of one day's change (remember we are looking at logs) will be $m = pa - (1-p)b = p(a+b) - b$, while the variance will be

$$\sigma^2 = pa^2 + (1-p)b^2 - (p(a+b) - b)^2 = pa^2 + b^2 - pb^2 - p^2(a+b)^2 + b^2 - 2bp(a+b) =$$
$$= p(1-p)(a+b)^2$$

Now, for large n, we can approximate $\frac{S_n - nm}{\sqrt{n}\sigma}$ with a standard normal. Since $S_n - S = \log \frac{Y_n}{Y}$, where Y is the initial price, multiplied by the up and down movements, we finally have that

$$\frac{\log \frac{Y_n}{Y} - n\left(pa - b\left(1-p\right)\right)}{\sqrt{np\left(1-p\right)}(a+b)} = Z$$

is (approximately, but well enough for n large) distributed like a standard normal.

For the purpose of our problem, we have $n = 1000, u = 1.012, d = .990, p = .52$. Hence $a = .0119, b = 0.101$. We wonder whether $\frac{Y_{1000}}{Y} \geq 1.3$, i.e., whether

$$Z \geq \frac{\log 1.3 - 1000\left(.52 \cdot .0119 - .101 \cdot .48\right)}{0.120\sqrt{1000 \cdot .52 \cdot .48}}$$

The probability of this event is, from the tables,

$$P[Z \geq -3.2149] = .99935$$

consistent with our previous calculation. Of course, the extensive use of logs causes a lot of uncontrollable rounding errors, and that accounts for the small discrepancies, as well as for the discrepancies with the solution manual.

Incidentally, the distribution of Y_n (the exponential of a normal RV) is known as a "lognormal" distribution, and, as already mentioned, is a classic favorite in theoretical financial models.

[3] Actually, this is true if we assume we can borrow money, in order to buy on credit, at 0 interest rate. The complete model (called the "binomial model", and used extensively for simulations) also calls for an interest rate greater than 0, and, consequently, the restrictions on u and d are such that we won't be able to gain no matter what, because credit would have a cost.

#31

You might recall the following fact: for any RV X, the number m that minimizes the function $\phi(m) = E\left[(X-m)^2\right]$ is $m = EX$. This is easily proved, by just expanding the expression, and observing that a parabola concave up is lowest at its vertex. This can be expressed by saying that "the number that minimizes the quadratic expected cost is the expected value", and provides a connection between the variance (and the idea of "quadratic cost"), and the expected value.

Another way of "measuring cost" is the one put forward here: minimize the expected ***absolute value*** of the error, instead of its square. Let's try to answer the question for a general continuous distribution (the argument for a discrete distribution is even more straightforward): let X have density f_X, and cdf F_X. We look for a number a that minimizes

$$\psi(a) = E[|X-a|] = \int_{-\infty}^{\infty} |x-a| f_X(x)\, dx$$

Due to the absolute value, there is no way we can solve this by deriving with respect to a - which illustrates why it is so much nicer to work with quadratic costs! By definition,

$$|x-a| = \begin{cases} x-a & x \geq a \\ a-x & x \leq a \end{cases}$$

so that

$$\psi(a) = \int_{-\infty}^{a} (a-x) f_X(x)\, dx + \int_{a}^{\infty} (x-a) f_X(x)\, dx =$$

$$= a\left(\int_{-\infty}^{a} f_X dx - \int_{a}^{\infty} f_X dx\right) - \left(\int_{-\infty}^{a} x f_X dx - \int_{a}^{\infty} x f_X dx\right)$$

This expression we can derive with respect to a in an easy way:

$$\psi'(a) = \int_{-\infty}^{a} f_X dx - \int_{a}^{\infty} f_X dx + 2a f_X(a) - 2a f_X(a) =$$

$$F_X(a) - (1 - F_X(a))$$

and this is zero when

$$2F_X(a) = 1$$

or

$$F_X(a) = P[X \leq a] = \frac{1}{2}$$

This value is called the "median" of the distribution (it represent the value for which X has probability 0.5 of being either lower or higher), and it is often used in describing empirical data (e.g., newspapers will report the "median price of housing", the "median income", etc. in a given area).

For our specific questions, we thus have that

a a is the median of the uniform distribution on $[0, A]$, and it is obvious that $a = \frac{A}{2}$

b a is the median of the exponential distribution, and this is the value such that

$$P[X \leq a] = P[X > a] = 0.5$$

i.e.

$$\begin{aligned} e^{-\lambda a} &= 0.5 \\ -\lambda a &= \log\left(\frac{1}{2}\right) \\ a &= \frac{\log(2)}{\lambda} = \tau \cdot \log(2) \end{aligned}$$

Note the connection between a and the expected value $\tau = \frac{1}{\lambda}$. It is the median that is often called "the half life" of a radioactive material, whose decay law is assumed to follow an exponential distribution. It is almost the same as the expected value, up to a factor of $\log(2) = .69315$.

Of course, we obtain the same result if we work directly with the given distributions, without referring to the general result.

a We look for a that minimizes

$$\psi(a) = \int_0^A |x-a| \frac{1}{A} dx = \frac{1}{A}\left(\int_0^a (a-x)\,dx + \int_a^A (x-a)\,dx\right) =$$

$$= \frac{1}{A}\left(a^2 - \frac{a^2}{2} + \left(\frac{A^2-a^2}{2}\right) - a\,(A-a)\right) = \frac{1}{A}\left(\frac{A^2}{2} - aA + a^2\right)$$

which, as a function of a, is a parabola, concave up, with a minimum at its vertex, i.e. at $a = -\frac{-A}{2} = \frac{A}{2}$

b We now look at minimizing

$$\int_0^\infty \lambda\,|x-a|\,e^{-\lambda x} dx = \lambda\left(\int_0^a (a-x)\,e^{-\lambda x} dx + \int_a^\infty (x-a)\,e^{-\lambda x} dx\right) =$$

$$= \lambda\left(a\left(\int_0^a e^{-\lambda x} dx - \int_a^\infty e^{-\lambda x} dx\right) - \left(\int_0^a xe^{-\lambda x} dx + \int_a^\infty xe^{-\lambda x} dx\right)\right)$$

Multiplying throughout the first factor λ, we get

$$a - \frac{1}{\lambda}\left(1 - 2e^{-\lambda a}\right)$$

(we have used the cdf of an exponential, $1 - e^{-\lambda x}$, as well as integrated by parts the terms of the form

$$\int_c^d xe^{-\lambda x} dx = \left[-\frac{1}{\lambda} xe^{-\lambda x}\right]_c^d + \frac{1}{\lambda}\int_c^d e^{-\lambda x} dx =$$

$$= \frac{1}{\lambda}\left(ce^{-\lambda c} - de^{-\lambda d}\right) + \frac{1}{\lambda^2}\left(e^{-\lambda c} - e^{-\lambda d}\right)$$

substituting appropriately).
This function of a will be minimized at $1 + 2e^{-\lambda a} = 1$, or $a = \frac{\log 2}{\lambda}$

#32 (C)

We only need to remember that for an exponential RV, with parameter λ

$$P[X > t] = e^{-\lambda t}$$

a We are asked to calculate

$$P[X > 2] = e^{-2\lambda} = e^{-2\cdot\frac{1}{2}} = e^{-1} = .36788$$

b We are now asked for

$$P[X > 10 \,|X > 9] = P[X > 9 + 1 \,|X > 9] = \frac{P[X > 10 \cap X > 9]}{P[X > 9]} = \frac{P[X > 10]}{P[X > 9]} = \frac{e^{-\lambda 10}}{e^{-\lambda 9}} =$$

$$= e^{-\lambda} = P[X > 1] = e^{-\frac{1}{2}} = .60653$$

Remark We have re-proved the remarkable property of the exponential distribution:

$$P[X > s + t \,|X > s] = P[X > t]$$

Remark Note that $\lambda$is the reciprocal of the expected value. Hence, when we measure time in hours, $\lambda$is measured in h^{-1}, and when we measure in minutes, it is measured in m^{-1}, etc. Hence, its value, in an application, will depend on the units we are using. The book should have specified that the value $\frac{1}{2}$ it was assigning to λ was in h^{-1}. Note, by the way, that the units used for λ are those of *frequencies.*

#34 (C)

Here, once we do the math, we get to see the difference between lifetimes that are exponentially distributed, and lifetimes that are not! Indeed, we already know (and the math would be the same as in the previous example), that, for the exponential case,

$$P[X > t + s \,|X > t] = P[X > s]$$

hence that

$$P[X > 10,000 + 20,000 \,|X > 10,000] = P[X > 20,000] = e^{-\frac{20000}{20000}} = e^{-1} = .36788$$

which is the same probability of going 20,000 miles that Smith originally had. If, however, the distribution is uniform on $[0, 40, 000]$, then

$$P[X > 10,000 + 20,000 \,|X > 10,000] = \frac{P[X > 30,000]}{P[X > 10,000]} = \frac{40,000 - 30,000}{40,000 - 10,000} = \frac{10,000}{30,000} = \frac{1}{3}$$

Of course, originally, the probability of getting 20,000 miles had been $\frac{1}{2}$.

Note In this case, X is measured in 1000s of miles - hence $\lambda$is measured in $(1000mi)^{-1}$

#36

A distribution with hazard function $h(t)$ is such that

$$P[X > t] = e^{-h(t)}$$

hence in our case,

a $P[X > 2] = e^{-\int_0^2 t^3 dt} = e^{-\frac{2^4}{4}} = e^{-4} = .018316$

b $P[.4 < X < 1.4] = e^{-\int_0^{0.4} t^3 dt} - e^{-\int_0^{1.4} t^3 dt} = e^{-\frac{.4^4}{4}} - e^{-\frac{1.4^4}{4}} = .61088$

c $P[X > 2\,|X > 1] = \frac{P[X>2]}{P[X>1]} = e^{-\int_1^2 t^3 dt} = e^{-4+\frac{1}{4}} = .023518$

#37

As usual, let's be careful with absolute values.

a $P\left[|X| > \frac{1}{2}\right] = P\left[X > \frac{1}{2}\right] + P\left[X < -\frac{1}{2}\right] = 2 \cdot \frac{1}{4} = \frac{1}{2}$

b To find the density, we can proceed in several ways. Let's, for instance, compute the cdf (obviously, $F_X = 0$ for $x < 0$, so we restrict to $x > 0$):

$$P[|X| \leq x] = P[-x \leq X \leq x] = \int_{-x}^{x} \frac{1}{2} dt = x$$

Hence, $|X|$ is uniform on $[0, 1]$.

#40 (C)

We work just as we did in the previous problem (note that $e^x > 0$ for any x, so the variable Y only takes positive values - we thus restrict to $y > 0$):

$$P\left[e^X \leq y\right] = P[X \leq \log y] = \log y$$

as long as $\log y \leq 1$, i.e., $y \leq e$ (which is obviously the top range for e^X), and $\log y \geq 0$, i.e. $y \geq 1$ (which is obviously the bottom range for e^X). The density of Y is the derivative of this function: always equal to 0, except for $1 \leq y \leq e$, where it is equal to $\frac{1}{y}$

Theoretical Exercises

#1

This is, really, a Calculus exercise. We are enforcing

$$\int_0^\infty ax^2 e^{-bx^2} = 1$$

The following method relies on the identity $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, which you may be familiar with, or not (it is proved on page 199 in the book), and the property of the Gamma function (see page 215): $\Gamma(x) = (x-1)\Gamma(x-1)$

Taking this identity for granted, we can use substitution, namely $x^2 = u$, to transform the integral into

$$\frac{a}{2}\int_0^\infty \frac{\sqrt{b}u^{\frac{1}{2}}}{\sqrt{b}} e^{-bu} du = \frac{a\Gamma\left(\frac{3}{2}\right)}{b\sqrt{b}}\int_0^\infty \frac{b(bu)^{\frac{3}{2}-1}}{\Gamma\left(\frac{3}{2}\right)} e^{-bu}$$

From the definition of the Gamma distribution (on page 237), we know that the integral is equal to 1 (read b as λ in the formula). Hence, we have

$$a\frac{\Gamma\left(\frac{3}{2}\right)}{b^{\frac{3}{2}}} = 1$$

which, using the fact that $\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$ results in

$$a = \frac{2b^{\frac{3}{2}}}{\sqrt{\pi}} = 2b\sqrt{\frac{b}{\pi}}$$

#2

Following the hint, we have (for absolutely continuous random variables)

$$E[Y] = \int_{-\infty}^\infty yf_Y(y)dy = \int_{-\infty}^0 yf_Y(y)dy + \int_0^\infty yf_Y(y)dy$$

Integrating by parts both integrals results in

$$[yF_Y(y)]_{-\infty}^\infty - \int_{-\infty}^\infty F_Y(y)dy$$

where F_Y is the cumulative distribution function of Y. To prove the assertions we only need to observe that the integral is precisely the right hand side of the equation (when a density function exists, the definition of cumulative distribution function in terms of < or $\leq$ makes no difference), and that, in order for the expected value to exist, the first term in the expression has to be zero (that is equivalent to the requirement that the integral $\int_{-\infty}^\infty yf_Y(y)dy$ exist!)

#3

The hint shows a pretty straightforward path. A less rigorous argument (that can, however, be extended to more general situations, once the necessary technicalities are settled) could go as follows.
Let X be a continuous random variable. We approximate it by X_n, where the range of X has been "cut up" in small segments $[a_k, a_{k+1})$, of length $\frac{1}{n}$, and $X_n = a_k$ for all outcomes such that $a_k \leq X < a_{k+1}$. We now have a discrete random variable, and we have, from Chapter 4

$$E[g(X_n)] = \sum_k g(a_k) P[X_n = a_k]$$

By the fact that $P[X_n = a_k] = \int_{a_k}^{a_{k+1}} f(x)dx$, and that the right hand side is now, essentially, a Riemann sum for our "target" integral, as $n \to \infty$, we have our proof.
Remark: the argument would work for *any* random variable X, if we had a rigorous method for passing to the limit in $\sum_k g(a_k) P[X \in [a_k, a_{k+1})]$ even when the distribution does not admit a continuous density. In fact, we would need to extend our definition of integral, in order for this to work, and that is precisely what the Lebesgue integration theory does for functions if a real variables, and general measure theory does for more general situations (recall that our X has an abstract space, the sample space, as its domain).

#5

Let's follow the hint:

$$E[X^n] = \int_0^\infty P[X^n > x^n] n x^{n-1} dx$$

which proves the statement.

#6

This problem has a simple answer: choose E_a as the event when X is *not equal* to a. It is not such a surprising example, as we are dealing with an *uncountable* intersection. In fact, if you take complements, the condition requires that the *union* of the complements of these sets (each having probability zero) have probability 1. This can obviously happen only because we are dealing with an uncountable collection of events.
This example reminded me of a different, more surprising, one, which you might feel has some relation to it (or maybe you won't feel that way - no matter: it is an interesting example, and it has big implications in unexpected areas, like celestial mechanics).
Consider the set of rational numbers, say, between 0 and 1. Since they are a countable set, we can write them up as a sequence, say, r_1, r_2, r_3, ... This is a *dense* set in [0,1]. Now, consider the set made up of the union of intervals of amplitude $\frac{\varepsilon}{2^n}$ around r_n, for a small ε. The total probability associated with these intervals is less than or equal to the sum of their amplitudes, which is equal to $\varepsilon \sum_n \frac{1}{2^n} = \varepsilon$. This is surprising, because ε can be as small as you wish, but the sets we are talking about are all open and dense, so that, in a topological sense, they are "big". The intersection over a sequence of such sets, as $\varepsilon \to 0$, will have probability zero. Topologically, though, it will be the intersection of a countable number of open dense sets - this

is called a “second category Baire set”, which, again, topologically is thought to be a “big” set (the intersection of second category Baire sets, is, again, a second category Baire set). The conclusion of this example is that “probabilistic” (or “measure theoretic”) size is very different from “”topological” size.

#10

We don’t need to be this lazy, but, thanks to a quick change of variables (it is a standard one: $x \mapsto \frac{x - EX}{\sqrt{\mathrm{Var}(X)}}$), we might as well prove that $f(x) = e^{-\frac{x^2}{2}}$ has an inflection point at $|x| = 1$. Since the second derivative of $f(x)$ is $(x^2 - 1)e^{-\frac{x^2}{2}}$ this is pretty obvious

#15

A simple way to show this is to evaluate

$$P[cX > x] = P[X > \tfrac{x}{c}] = e^{-\lambda \frac{x}{c}} = e^{-\frac{\lambda}{c}x}$$

#16

The hazard rate for an absolutely continuous random variable *X* is defined as the density formally defined by $\frac{P[X \in dx \mid X > x]}{dx}$, which, with a little care, can be seen to translate, in the case of absolutely continuous random variables, into $\frac{f_X(x)}{1 - F_X(x)}$, as discussed on page 234 (section 5.5.1). It is a very intuitive tool in survival and reliability theory. For a uniform variable on $[0, a]$ (notice that, for distributions with a continuous density, it makes no difference if we consider open, closed, semi-open, etc. intervals), this translates to

$$\frac{\frac{1}{a}}{\frac{a-x}{a}} = \frac{1}{a - x}$$

Not surprisingly, it explodes as *x* approaches *a*: after all our gadget *has to fail* at some point between 0 and *a*, and as we approach the end of the interval “time for failure is running out”!

#23

The density of a Weibull distribution can be defined from

$$R(t) = P[X > t] = e^{-at^b}$$

(In reliability/survival analysis we will always have $t \geq 0$). While there are approximating argument to justify this choice, the most obvious reason is seen by comparing with the exponential distribution, where

$$P[X > t] = e^{-at}$$

As the question implies, choosing values of *b* less than, equal to, to greater than 1, corresponds to having hazard rates that are decreasing (a “rejuvenation” effect), constant (“memoryless”), or increasing (“aging” effect). Thus, the Weibull family provides an easy toolbox for a quick and dirty description of the different age-related scenarios for the likelihood of an upcoming failure. The hazard rate calculation is an exercise in derivation:

$$\frac{f_X(t)}{R(t)} = \frac{-R'(t)}{R(t)} = -\frac{d\log R(t)}{dt} = -\frac{d\left(-at^b\right)}{dt} = abt^{b-1}$$

The conclusion about the different behavior as $b > 1$, $b = 1$, $b < 1$ follows now easily

#27

The cumulative distribution function for X is $P[X \leq x] = \frac{x-a}{b-a}$. Let's define $Y = \alpha X + \beta$. Since $\alpha a + \beta \leq Y \leq \alpha b + \beta$, we need to make sure that $\alpha a + \beta = 0$, and $\alpha b + \beta = 1$. Thus, we want $\alpha = \frac{1}{b-a}$, $\quad \beta = \frac{a}{a-b}$. Checking, for example, the cumulative distribution function will confirm this is he right choice:

$$P[Y \leq y] = P\left[X \leq \frac{y-\beta}{\alpha}\right] = P[X \leq a - (a-b)y] = y$$

(for $0 \leq y \leq 1$), by directly substituting.

#29

Note that F is a monotone increasing function. Since we are assuming that it is *strictly* monotone increasing function (and, one could be careful and extend the argument to the non strictly monotone case), it has a proper inverse F^{-1} (in the non strict case, one can use the "right inverse"). Then,

$$P[F(X) \leq y] = P\left[X \leq F^{-1}(y)\right] = F\left(F^{-1}(y)\right) = y$$

This observation is sometimes used to simulate a random variable with a given distribution function, starting from a program that simulates a uniform random variable: starting from a uniform random variable Y, the random variable $X = F^{-1}(Y)$ has distribution function F. This method is easy, whenever F has an easy inverse e.g., for exponential variables, but not for normal variables), but has its own numerical pitfalls - for example, simulating an exponential variable requires the computation of a logarithm, which is computer intensive, as well as less precise than algebraic calculations.

#31

We can compute the probability density function in various ways. A plain route is the following:

$$P\left[e^X \leq y\right] = P[X \leq \log y] = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\log y} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

Taking the derivative with respect to y, we obtain

$$f_Y(y) = \frac{e^{-\frac{(\log y-\mu)^2}{2\sigma^2}}}{y\sqrt{2\pi\sigma^2}}$$

This choice for stock price modeling has the advantage of simplicity, since the logarithm of the random variable has a convenient normal distribution, but it does have a number of drawbacks. One that got a lot of attention in the wake of the recent financial crisis is its lack of "fat tails" (outliers are very unlikely), but there are other concerns as well as to its appropriateness.