Question 1

a) $E(S_{it}) = E(\alpha A_i + (1-\alpha)u_{it})$

$= \alpha A_i + (1-\alpha)E(u_{it})$

$= \underline{\alpha A_i}$ given $E(u_{it}) = 0$.

b) $Var(S_{it}) = E((S_{it} - E(S_{it}))^2$

$= E(\alpha A_i + (1-\alpha)u_{it} - \alpha A_i)^2$

$= E((1-\alpha)u_{it})^2$

$= (1-\alpha)^2 E(u_{it})^2$

Now we are given that $Var(u_{it}) = \sigma^2$

But

$Var(u_{it}) = E(u_{it} - E(u_{it}))^2$

$= E(u_{it})^2$

Therefore, we can substitute σ^2 in for $E(u_{it})^2$, as

$\underline{Var(S_{it}) = (1-\alpha)^2\sigma^2}$.

c) Consider setting a single test. If $\alpha = 0$, then the test results are completely uninformative: the score is purely random and setting a test would be a waste of resources. At the other extreme, if $\alpha = 1$, then the test is fully revealing of underlying ability.

Question 2

a) What is the variance of the sum of scores taken by a given student?

Given that each student faces the same noise distribution, we can focus on student 1, and ask: what is $Var(T_{1i} + T_{1j})$, where i and j index two separate tests?

(The answer will be the same for anyother student.)

Now $T_{1i} = A_1 + u_{1i}$, and

$$E(T_{1i}) = A_1 + E(u_{1i}) = A_1 + E(u_i) = A_1,$$ given that the error distribution is not student-specific.

Also $T_{1j} = A_1 + u_{1j}$, and $E(T_{1j}) = A_1$.

Thus,

$$Var(T_{1i} + T_{1j}) = Var(A_1 + u_{1i} + A_1 + u_{1j})$$
$$= Var(A_1 + A_1 + u_i + u_j)$$
$$= Var(u_i + u_j)$$

Now, given the long derivation in the 'Properties' note, we know that

$$Var(u_i + u_j) = Var(u_i) + Var(u_j) + 2Cov(u_i, u_j).$$

Here, given the random draw on one test is unrelated to the random draw on any other test, we have

$$Var(u_i + u_j) = Var(u_i) + Var(u_j) = 2\sigma^2$$

b) Let $\bar{T}_1$ be the sample average score for student 1, computed over N tests.

Question: what is $Var(\bar{T}_1)$?

$$Var(\bar{T}_1) = Var\left(\frac{1}{N}\sum_{i=1}^{N} T_{1i}\right)$$

$$= \frac{1}{N^2} Var\left(\sum_i T_{1i}\right)$$

$$= \frac{1}{N^2} \sum_i Var(T_{1i})$$

$$= \frac{1}{N^2} \sum_i Var(u_i) \quad \text{, given independence.}$$

$$= \frac{1}{N^2} \sum_i \sigma^2$$

$$= \frac{1}{N} \sigma^2 \;//$$

The same result holds for any student, given the assumptions made about the 'iid' noise distribution.

c) Question: what is the limit of the variance of the student average test score as N, the number of tests, gets arbitrarily large?

$\therefore$ Consider $\lim_{N \to \infty} \left(\frac{1}{N} \sigma^2 \right)$

Clearly, σ^2 is a fixed number, assumed to be finite.

Then as N gets very large, so $\frac{1}{N}\sigma^2$ will tend to zero.

Thus we write

$$\lim_{N \to \infty} \left(\frac{1}{N} \sigma^2 \right) = 0. //$$

Note the significance of this. If N is very large, then our student average test score will exhibit <u>zero</u> variation.

In turn, the <u>sample</u> average score will be equal to the student's underlying ability.

Thus, for large N, we have a way of recovering unobserved ability.

d) With very many tests, the school could estimate the underlying ability of each student by averaging the test scores of that student. In part c), we saw that the 'noise' component becomes inconsequential as N gets large, and the sample average score serves as an unbiased estimator of fixed underlying ability as N gets big. (Another way of saying this would be that the sample average is a <u>consistent</u> estimator of underlying student ability.)

Question 3

a) Note that it is possible that student 1 could get a very high draw on the noise term (of the order of $+1$) at the same time as student 2 received a very low draw (in the region of -1). That being the case, then given the underlying ability distribution — here, $A_1 = 8.1$ and $A_2 = 10$ — student 1 could actually receive a higher score on the single test, and so be admitted. This would be at odds with the school's objective, but it is clearly a risk that arises in setting just one test.

[Aside: in the Bonus question, you are asked to compute the precise probability that student 1 receives a higher score than student 2. (I include this more for your interest.)]

b) With a more accurate test — that is, one with a lower variance — the problem arising in the previous part can be avoided entirely, as long as the variance is sufficiently low. [The next part is rather discursive — I provide a more succinct answer later.]

Specifically, the variance of a uniform random variable $u_i \sim U(a,b)$ is given by

$$Var(u_i) = \frac{1}{12}(b-a)^2.$$

Now, if the distribution is symmetric around zero, let $x = b$ and $-x = a$, and we can write

$$Var(u_i) = \frac{1}{12}(x-(-x))^2 = \frac{1}{12}(2x)^2$$

$$= \frac{1}{3}x^2$$

Thus, in this less general symmetric about zero case, where $u_i \sim U(-x, x)$, we can see that reducing x — in this instance, below 1 — lowers the variance, as we would expect.

Now, given the 'support' of the distribution is bounded (between $-x$ and x), we can rule out any score overlap as long as $x-(-x) = 2x < 1.9$, so x needs to be less than $0.8 < 1$. That being the case, any chance of overlap can be ruled out entirely.

All that said, we can get to the short concise answer:

"Using a lower variance test, with an associated random variable $u_i \sim U(-x, x)$, the problem of the lower ability student being admitted will still arise if $0.8 \leq x < 1$, but will be avoided if $x < 0.8$. That way, in the latter case, the error terms could never be spread out enough to reverse the effects of the fixed ability gap.

c) With a pass threshold of 9, we can compute the probability that each candidate will pass, given $u_i \sim U(-1, 1)$, as follows:

$$\Pr(T_1 \geqslant 9) = \Pr(A_1 + u_1 \geqslant 9) = \Pr(u_1 \geqslant .9),$$

given $A_1 = 8.1$.

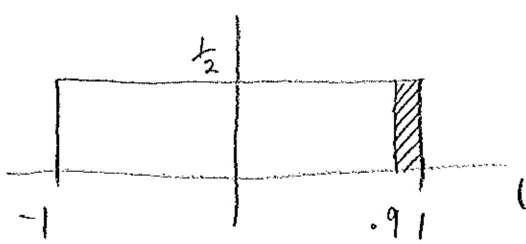
Now $\Pr(u_1 \geqslant 0.9) = \int_{.9}^{1} f(u_1)\, du_1$

$$= \int_{.9}^{1} \frac{1}{2}\, du_1 = \frac{1}{2}\Big[u_1\Big]_{.9}^{1}$$

$$= \frac{1}{2} \times \frac{1}{10} = \frac{1}{20} //$$

In other words, student 1 has only a 5-percent chance.

Aside: all we are doing here is calculating the shaded area below.



That area $= (1 - 0.9) \times \frac{1}{2} = \frac{1}{20}$.

We can also compute it by evaluating the above integral.

The corresponding probability for student 2 is given by

$$\Pr(T_2 \geqslant 9) = \Pr(A_2 + u_2 \geqslant 9) = \Pr(u_2 \geqslant -1),$$

noting that $A_2 = 10$.

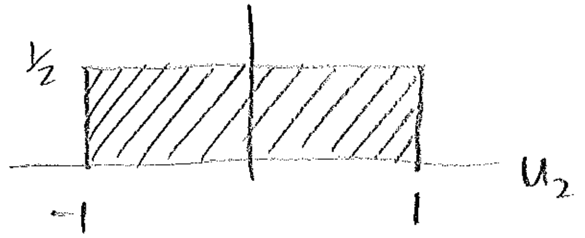
The formal approach involves computing

$$\int_{-1}^{1} f(u_2)\, du_2 = \int_{-1}^{1} \frac{1}{2}\, du_2 = \frac{1}{2}\Big[u_2\Big]_{-1}^{1}$$

$$= \frac{1}{2}\left[1-(-1)\right] = 1 /\!/$$

Thus, there is a 100 percent chance that student 2 is guaranteed to pass the test — that is, score $\geqslant 9$.

We can see the same by computing the shaded area below: it must have an area = 1, in virtue of being a proper density.

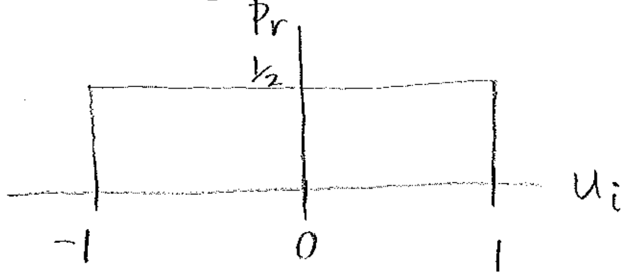


d) <u>What do you get?</u>

Solution to Bonus Question

We can work out the probability that student 1 earns a higher score than student 2 in this context, as follows:

Start with the distribution of the test score 'shocks' for each student i, given by $u_i \sim U(-1, 1)$, so the probability density function (p.d.f.) can be sketched as:



Question: what is $Pr(T_1 > T_2)$?

We have that $T_i = A_i + u_i$.

$$\text{So } Pr(T_1 > T_2) = Pr(A_1 + u_1 > A_2 + u_2)$$
$$= Pr(u_1 - u_2 > A_2 - A_1)$$
$$= Pr(u_1 - u_2 > 1.9) \text{ or}$$
$$= Pr(u_1 > 1.9 + u_2).$$

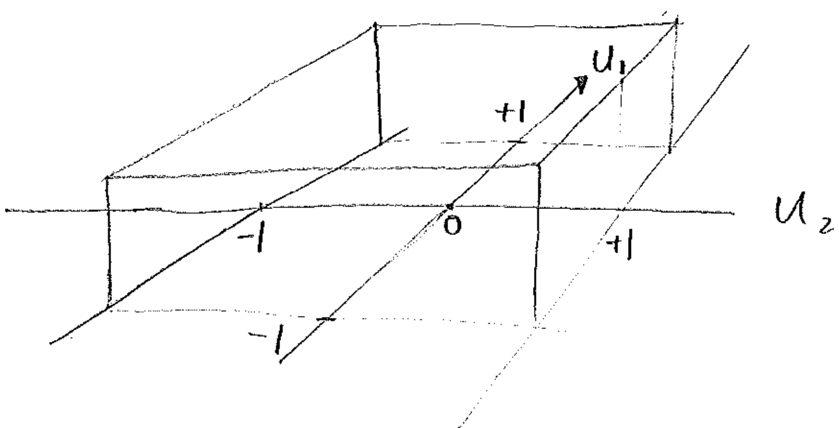
Now it turns out that we can draw a simple picture of this 'event'.

Start by noting that both u_1 and u_2 are assumed to be independently and identically distributed ('iid').

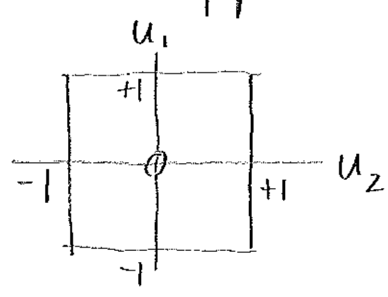
That implies that the particular value that u_1 takes on is completely uninfluenced by the value that u_2 takes on.

For instance, if $u_2 = -0.82$, then u_1 will still be drawn randomly from the $[-1, 1]$ interval, according to a uniform distribution. And the same applies if $u_2 = 0.569$ etc.

Thus, as we move along the u_2 axis, we have a series of uniform distributions that govern u_1.



The picture above is actually a drawing of the <u>joint</u> distribution of (u_1, u_2). Clearly, this is defined for $-1 \leq u_1, u_2 \leq +1$, giving the 'support' of the distribution as:



The joint distribution is clearly going to be uniformly distributed above that square base, given our independence assumption – see prior reasoning.

So the question is: what height will it have? We work this out in a way analogous to the method we used for a univariate (i.e. one-dimensional) uniform distribution.

The base has an area of $2 \times 2 = 4$. But the volume of the 'box' has to equal 1 for it to represent a proper joint density. So we have

$$\text{volume} = \text{base} \times \text{height}$$

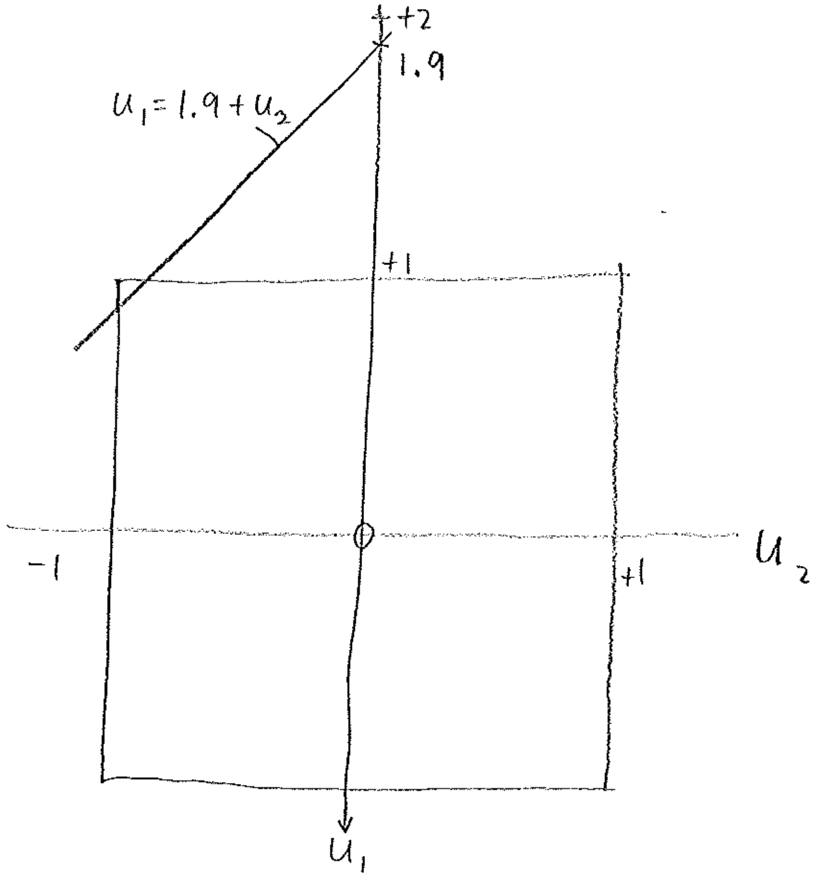
$$\Rightarrow \quad 1 = 4 \times \text{height}$$

$$\Rightarrow \quad \text{height} = 1/4.$$

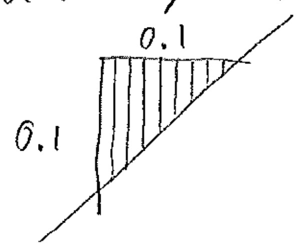
Now we are in a good position to calculate $\Pr(u_1 > 1.9 + u_2)$. Let us take an arial view of the base, and plot the curve

$$u_1 = 1.9 + u_2$$

on this, yielding:



Now focus on the small triangle: the area to north-west, which is shaded below, captures those



combinations of values of u_1 and u_2 for which $u_1 > 1.9 + u_2$, if we ignore the line itself.

What is the area of this base? It is $0.1 \times 0.1 \times 0.5 = 0.005$ But we know the height of the 'box' $= \frac{1}{4} = 0.25$. Therefore, the volume of the whole shape $= 0.005 \times 0.25 = \underline{0.00125}$.

This is the relevant probability: a quarter of one percent, which is a very small number.

We can be more formal about the preceding, though this takes us into the realm of joint distributions (beyond the scope of what I will cover). In that realm, we can write the joint p.d.f. as

$$f(u_1, u_2).$$

Now, under independence, this is equal to the product of what are called the 'conditional density function of u_1, given u_2' and the marginal density of u_2. Thus

$$f(u_1, u_2) = f(u_1 | u_2) f(u_2).$$

This follows from the definition of the conditional density:

$$f(u_1 | u_2) = \frac{f(u_1, u_2)}{f(u_2)}$$

simply be re-arranging.

But consider the density of u_1 given u_2. Under independence, the density of u_1 is completely unaffected by the value that u_2 takes on. Thus, $f(u_1 | u_2) = f(u_1)$.

But then

$$f(u_1, u_2) = f(u_1) f(u_2) = \frac{1}{2} \times \frac{1}{2} = 1/4. \quad //$$

Bonus question, part b)

In part a), we calculated the probability that student 1 scored higher than student 2 on a single test. For now, denote that probability p.

Under the assumption that a student's performance on one test is unrelated to ('independent of') her performance on any other test, then the probability of the event that ($T_{11} > T_{21}$ and $T_{12} > T_{22}$) is equal to the probability that ($T_{11} > T_{21}$) multiplied by the probability that ($T_{12} > T_{22}$). Under the independence assumption, those probabilities are equal (to $p = 0.00125$). Thus the answer is given by $p \times p = (0.00125)^2 = 0.00000156$

This is a <u>very</u> small number indeed!

c) Note that, given two tests, we can also compute the probability that student 1 comes first on at least one of them

Let us construct the full set of possibilities:

	Test 1	Test 2
student 1 first	p	p
student 1 2nd	$(1-p)$	$(1-p)$

[Please refer over the page for the full set of possible events and their associated probabilities.]

Thus, the probability that student 1 is first on both tests is (p^2), as in part b).

The probability that student 1 is first on just one test is given by $p \times (1-p) + (1-p) \times p$ — that is, the probability that student 1 is first on Test 1 plus the probability that student 1 is first on Test 2.

Thus, the probability that student 1 is first on at least one test is given by

$$p^2 + 2p(1-p),$$

and we can substitute in our value of p.

Just to check, the complete set of events can be listed as follows:

Student 1's position on each test		Corresponding probability of the joint event
1	1	$p \times p$
1	2	$p \times 1-p$
2	1	$1-p \times p$
2	2	$1-p \times 1-p$

Given these events are exhaustive and mutually exclusive (so we rule out ties), if we sum the right-most column, it should sum to 1:

$$p^2 + 2p(1-p) + (1-p)^2$$

$$= p^2 + 2p - 2p^2 + 1 - 2p + p^2 = 1 //.$$