Krzys' Ostaszewski http://www.math.ilstu.edu/krzysio/
Author of the BTDT Manual for Course P/1
available at http://smartURL.it/krzysioP or http://smartURL.it/krzysioPe
Instructor for online Course P/1 seminar: http://smartURL.it/onlineactuary
Exercise for May 14, 2005

X and Y are independent and both distributed uniformly from 0 to 20. Find the probability density function of $Z = 25X - 10Y$.

A. $f_Z(z) = 1.5$ where non-zero

B. Stepwise formula:

$$f_Z(z) = \begin{cases} \dfrac{200 - z}{100000}, & -200 \le z < 0, \\ \dfrac{1}{300}, & 0 \le z < 300, \\ \dfrac{500 + z}{100000}, & 300 \le z \le 500. \end{cases}$$

C. $f_Z(z) = \dfrac{1}{200}$ where non-zero

D. $f_Z(z) = 200e^{-\frac{1}{200}z}$, for $z > 0$, zero otherwise

E. Stepwise formula:

$$f_Z(z) = \begin{cases} \dfrac{200 + z}{100000}, & -200 \le z < 0, \\ \dfrac{1}{500}, & 0 \le z < 300, \\ \dfrac{500 - z}{100000}, & 300 \le z \le 500. \end{cases}$$

Solution.

This problem can be solved in a simplified way, but we will discuss three possible solutions, in a drawn-out fashion, in order to fully explain approached to such problems involving sums or differences of random variables. To begin with, note the following:

$$f_X(x) = \frac{1}{20} \quad \text{for } 0 < x < 20, \text{ and zero otherwise,}$$

$$f_Y(y) = \frac{1}{20} \quad \text{for } 0 < y < 20, \text{ and zero otherwise,}$$

and

$$f_{X,Y}(x, y) = \frac{1}{20} \cdot \frac{1}{20} = \frac{1}{400} \quad \text{for } 0 < x < 20 \text{ and } 0 < y < 20, \text{ and zero otherwise.}$$

There are actually three possible approaches to solving this: the multivariate transformation approach, the convolution approach, and the CDF approach. You actually should know all of them for the exam P, so all three will be presented here. They are related and their conclusions can be derived from each other.

• *Multivariate transformation*

Let $W = 10Y$ and $Z = 25X - 10Y$. This defines a transformation

$$(W,Z) = \Phi(X,Y) = (10Y, 25X - 10Y),$$

whose inverse is

$$\Phi^{-1}(W,Z) = \left(\frac{W+Z}{25}, \frac{W}{10}\right).$$

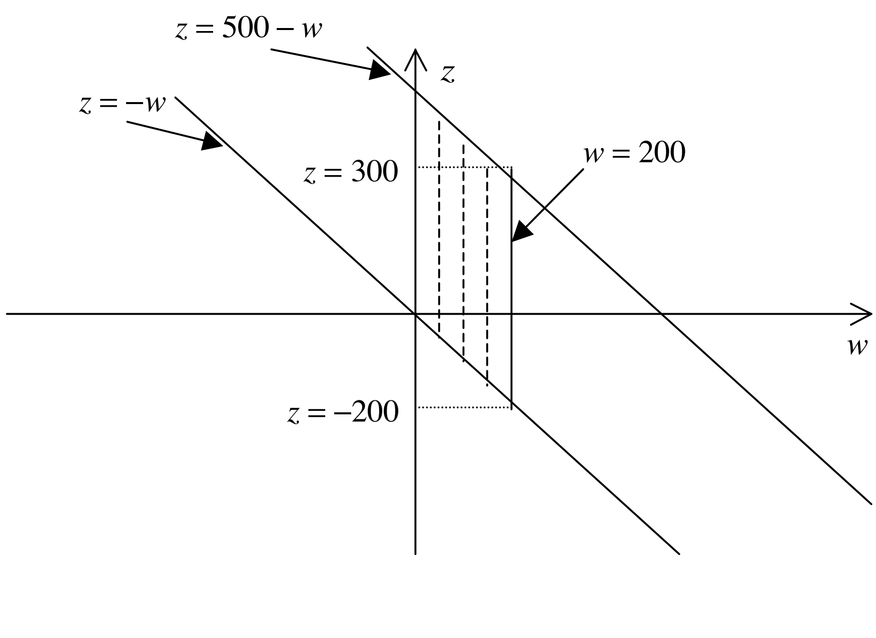
You might wonder: how do I know that I am supposed to pick $W = 10Y$? In fact, you are not supposed to pick anything. Your objective is to find a second function of X and Y such that you will be able to find the inverse of the transformation so obtained. There is no unique answer. In this case, $W = Y$ would do the job, so would $W = 25X + 10Y$, and so would inifinitely many other choices. The key insight is that you must be able to quickly find Φ^{-1} and then find the determinant of its derivative (the *Jacobian)*. Let us find that derivative now (we switch to lower case variables because this is what we will use in the density calculation):

$$\left(\Phi^{-1}\right)'(w,z) = \begin{bmatrix} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial w} & \frac{\partial y}{\partial z} \end{bmatrix} = \begin{bmatrix} \frac{1}{25} & \frac{1}{25} \\ \frac{1}{10} & 0 \end{bmatrix}.$$

Therefore,

$$\frac{\partial(x,y)}{\partial(w,z)} = \det \begin{bmatrix} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial w} & \frac{\partial y}{\partial z} \end{bmatrix} = -\frac{1}{250}.$$

This gives $f_{W,Z}(w,z) = f_{X,Y}\left(x(w,z), y(w,z)\right) \cdot \left|\frac{\partial(x,y)}{\partial(w,z)}\right| = \frac{1}{400} \cdot \frac{1}{250} = \frac{1}{100000}$. We also have to figure out the ranges for w and z. As $w = 10y$ and $0 < y < 20$, we have $0 < w < 200$. Also, as $0 < x < 20$ and $0 < y < 20$, and $z = 25x - 10y$, we have $-w < z < 500 - w$ (as well as $-z < w < 500 - z$). Graphically,



Given that, we can now figure out the marginal density of Z:

$$f_Z(z) = \int\limits_{\text{all values of } w} f_{W,Z}(w,z)\,dw =$$

$$= \begin{cases} \displaystyle\int_{-z}^{200} \frac{1}{100000}\,dw, & -200 \le z < 0, \\ \displaystyle\int_{0}^{200} \frac{1}{100000}\,dw, & 0 \le z < 300, \\ \displaystyle\int_{0}^{500-z} \frac{1}{100000}\,dw, & 300 \le z \le 500. \end{cases} = \begin{cases} \dfrac{200+z}{100000}, & -200 \le z < 0, \\ \dfrac{1}{500}, & 0 \le z < 300, \\ \dfrac{500-z}{100000}, & 300 \le z \le 500. \end{cases}$$

Answer E.

• *The convolution method*

Recall that if X and Y have a continuous joint distribution and are continuous, then the density of $X+Y$ is

$$f_{X+Y}(s) = \int_{-\infty}^{+\infty} f_{X,Y}(x, s-x)\,dx.$$

If X and Y are independent, then

$$f_{X+Y}(s) = \int_{-\infty}^{+\infty} f_X(x) f_Y(s-x)\,dx.$$

In this case, we are adding

• $U = 25X$, which has the uniform distribution on (0, 500), and is independent of

• $V = -10Y$, which has the uniform distribution on (–200, 0).

The density of U is $\frac{1}{500}$, where non-zero, and the density of V is $\frac{1}{200}$, where non-zero.

Thus

$$f_{U+V}(s) = \int_{-\infty}^{+\infty} f_U(u)\cdot f_V(s-u)\,du = \int\limits_{\substack{0\le u\le 500 \text{ AND} \\ -200\le s-u\le 0}} \frac{1}{500}\cdot\frac{1}{200}\,du = \int\limits_{\substack{0\le u\le 500 \text{ AND} \\ s\le u\le s+200}} \frac{1}{500}\cdot\frac{1}{200}\,du =$$

$$= \begin{cases} \displaystyle\int_{0}^{s+200} \frac{1}{100000}\,du, & -200 \le s < 0, \\ \displaystyle\int_{s}^{s+200} \frac{1}{100000}\,du, & 0 \le s < 300, \\ \displaystyle\int_{s}^{500} \frac{1}{100000}\,du, & 300 \le s \le 500. \end{cases} = \begin{cases} \dfrac{200+s}{100000}, & -200 \le s < 0, \\ \dfrac{1}{500}, & 0 \le s < 300, \\ \dfrac{500-s}{100000}, & 300 \le s \le 500. \end{cases}$$

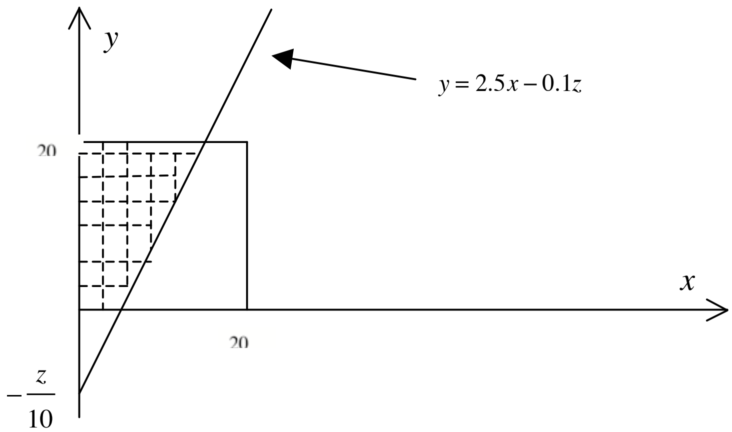
Answer E, again.

• *The CDF method*

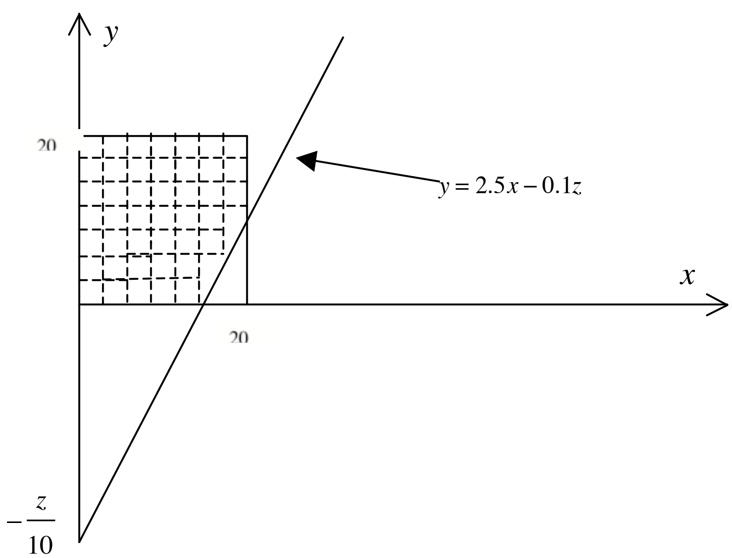
We have $Z = 25X - 10Y$. Let us find the CDF of it directly. We have:

$$F_Z(z) = \Pr(Z \le z) = \Pr(25X - 10Y \le z).$$

This probability can be obtained by simply taking the integral of the joint density of X and Y over the region where $25X - 10Y \le z$. The figure below shows the region where the joint density is allowed to play, and the shaded area is where $25X - 10Y \le z$.



The problem is that the position of the line $y = 2.5x - 0.1z$ can vary as z varies, and we get different results in different cases. The line crosses the point $(20,0)$ when $z = 500$. If $z \ge 500$ then the line does not go through the 20 by twenty square at all and the probability of being above the line is 1. Since the slope of the line is more than 1, the next point where a change occurs is when the line crosses the point (20,20), which occurs for $z = 300$. The value of the CDF of Z for any z between 300 and 500 is the area shown in this figure:



The point where the line crosses the x axis is at $(x, y) = \left(\frac{z}{25}, 0\right)$ and the point where it crosses the line $x = 20$ is at $(x, y) = \left(20, 50 - \frac{z}{10}\right)$. The area of the bottom-right triangle left out of the calculation of probability is therefore $\frac{1}{2}\cdot\left(20-\frac{z}{25}\right)\cdot\left(50-\frac{z}{10}\right)$, and the corresponding probability is $\frac{1}{400}$ times that, so that the probability we are looking for is:

$$F_Z(z)=1-\frac{1}{800}\cdot\left(20-\frac{z}{25}\right)\cdot\left(50-\frac{z}{10}\right).$$

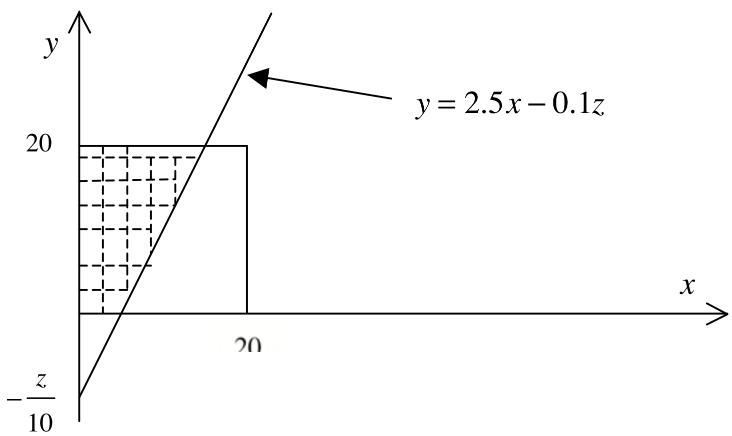
The corresponding density is:

$$f_Z(z)=F_Z'(z)=-\frac{1}{800}\cdot\left(-\frac{1}{25}\right)\cdot\left(50-\frac{z}{10}\right)-\frac{1}{800}\cdot\left(20-\frac{z}{25}\right)\cdot\left(-\frac{1}{10}\right)=$$

$$=\frac{1}{400}-\frac{z}{800\cdot 25\cdot 10}+\frac{1}{400}-\frac{z}{800\cdot 25\cdot 10}=\frac{500-z}{100000},$$

for $300\le z\le 500$. This is the same answer in this range of values of z that we obtained before.

The second case starts with z for which the point (20, 20) is crossed by the line, i.e., z = 300. This case is generally described by this figure:



This case ends when the line crosses the origin, i.e., when $z=0$. Between $z=0$ and z = 300, the probability we want to find is just the area of the shaded parallelogram as a fraction of the area of the 20 by 20 square. The bottom side of the parallelogram has length $\frac{z}{25}$ (from y = 0 and the equation of the line) and the top side has length $8+\frac{z}{25}$ (from y = 20 and the equation of the line), so that its area is $20\cdot\left(4+\frac{z}{25}\right)=80+\frac{4}{5}z.$ As a fraction of the entire area, this is

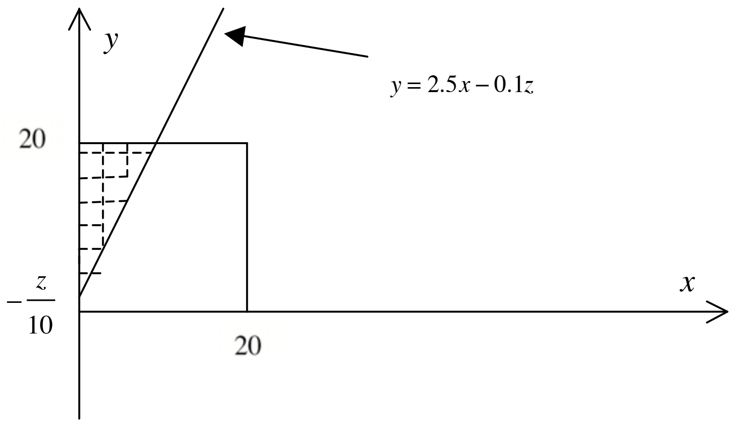
$$F_Z(z)=\frac{80+\frac{4}{5}z}{400}=\frac{1}{5}+\frac{z}{500}.$$

Therefore,

$$f_Z(z)=F_Z'(z)=\frac{1}{500}\quad\text{for } 0\le z\le 300.$$

Again, this is the same answer we obtained before.

The final case is when the line crosses the *y*-axis above the origin, but below 20. When the line crosses the origin, $z = 0$. When the line crosses the *y*-axis at the point (0, 20), we have $z = -200$. In this case, the figure looks as follows:



When $y = 20$, we have $x = 8 + \frac{z}{25}$, so that the area of the marked triangle is

$$\frac{1}{2}\cdot\left(8+\frac{z}{25}\right)\cdot\left(20+\frac{z}{10}\right).$$

The 20 by 20 square has the area of 400, so that

$$F_Z(z) = \frac{\frac{1}{2}\cdot\left(8+\frac{z}{25}\right)\cdot\left(20+\frac{z}{10}\right)}{400} = \frac{\left(8+\frac{z}{25}\right)\cdot\left(20+\frac{z}{10}\right)}{800}.$$

Therefore,

$$f_Z(z) = F_Z'(z) = \frac{1}{800}\left(\frac{1}{25}\left(20+\frac{z}{10}\right)+\frac{1}{10}\left(8+\frac{z}{25}\right)\right) = \frac{1.6+\frac{z}{125}}{800} = \frac{200+z}{100000},$$

for $-200 \le z \le 0$. This is again the same formula we obtained before. Answer E again.