Solutions of problems.

First of all, I would like to speak about multidimensional normal (Gaussian) distributions.

But first, about one-dimensional normal distributions.

For a pair of real numbers (a, b), $b > 0$, we say that a random variable X has normal distribution with parameters (a, b) if it has a continuous distribution with density

$$p_{a,\,b}(x) = \frac{1}{\sqrt{2\pi b}}\, e^{-(x-a)^2/2b}. \tag{1}$$

It turns out that these parameters have the following meaning: the mean $a = E(X)$ and the variance $b = \mathrm{Var}(X)$.

The parameters for an n-dimensional normal distribution are: an n-dimensional vector $\boldsymbol{a}$ and a symmetric $n \times n$ matrix B. If the matrix B is positive definite (and therefore nonsingular), there exists the inverse (symmetric, positive-definite) matrix $B^{-1} = Q = (q_{ij})$, and we can define the n-dimensional normal distribution: the n-dimensional random vector $\boldsymbol{X}$ with components X_1, ..., X_n has the n-dimensional normal distribution with parameters $(\boldsymbol{a}, B)$ if it has a continuous n-dimensional distribution with density

$$p_{\boldsymbol{a},\,B}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(B)}}\ \exp\{-(\boldsymbol{x}-\boldsymbol{a})^T\, B^{-1}\,(\boldsymbol{x}-\boldsymbol{a})/2\}, \tag{2}$$

where the vectors are written as column ones; ^{T} denotes the transposed, so that $\boldsymbol{x}^T$ and $\boldsymbol{a}^T$ are row vectors. In the coordinates, we can rewrite this formula and reformulate the statement as: the joint distribution of n random variables X_1, ..., X_n is, by definition, normal with parameters $\big(a_1, ..., a_n;\ (b_{ij})_{i,\,j=1}^n\big)$ if their joint probability density is

$$p_{a_1,\,...,\,a_n;\,(b_{ij})}(x_1, ..., x_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(B)}}\ \exp\Big\{-\sum_{i,\,j=1}^{n} q_{ij}\,(x_i - a_i)(x_j - a_j)/2\Big\}. \tag{3}$$

It takes some effort proving that the integral of the density (2), (3) over $\mathbb{R}^n$ is equal to 1. The meaning of the parameters: $\boldsymbol{a}$ is the vector of the expectations:

$$a_i = E(X_i), \tag{4}$$

and $B = (b_{ij})$ is the covariance matrix:

$$b_{ij} = \mathrm{Cov}(X_i, X_j). \tag{5}$$

In particular, the diagonal entries of the matrix B are the variances $\mathrm{Var}(X_i)$.

If all off-diagonal $b_{ij} = 0$ (i. e., the random variables X_1, ..., X_n are *uncorrelated*), B is the diagonal matrix with diagonal entries $\mathrm{Var}(X_i) = \sigma_i^2$, the matrix $Q = B^{-1}$ is also diagonal, with diagonal entries σ_i^{-2}, and formula (3) turns to

$$p_{X_1,\,...,\,X_n}(x_1, ..., x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\cdot\sigma_i}\, e^{-(x_i-a_i)^2/2\sigma_i^2}; \tag{6}$$

which means that the components X_i are mutually independent and normal with parameters (a_i, σ_i^2).

With some calculations, one can prove that if the collection of n random variables has the normal distribution with density (3), then each component X_i taken separately has a one-dimensional normal distribution (with parameters (a_i, b_{ii}), of course; and every subcollection X_{i_1}, ..., X_{i_k} of this collection has a k-dimensional normal distribution.

So, now: a control question: X has a normal distribution with parameters $(-1, 1)$, and Y a normal distribution with parameters $(3, 5)$; and $\mathrm{Cov}(X, Y) = 2$. What is the joint distribution of X and Y?

If you answer: two-dimensional normal, with parameters $\left(-1,\ 3;\begin{pmatrix}1 & 2\\ 2 & 5\end{pmatrix}\right)$, you are wrong: not necessarily (but, of course, *possible*). The things we talked above where all *under the assumption that* X_1, ..., X_n *have a joint normal distribution*; and we know only something about the individual distributions of X, Y, but nothing about their joint distribution.

Another question: X and Y have normal distributions with parameters, correspondingly, $(-1, 1$ and $(3, 5)$; and these random variables are *independent.* What is the joint distribution of X, Y?

This time, we do know something about the joint distribution of our random variables: their *independence* is just about it. We have that the joint density is given by

$$\begin{aligned} p_{X,\,Y}(x, y) = p_X(x)\cdot p_Y(y) &= \frac{1}{\sqrt{2\pi}}\, e^{-(x+1)^2/2}\cdot\frac{1}{\sqrt{2\pi\cdot 5}}\, e^{-(y-3)^2/2\cdot 5} \\ &= \frac{1}{2\pi\sqrt{5}}\, \exp\{-[(x+1)^2/2 + (y-3)^2/2\cdot 5]/2\}, \end{aligned} \tag{7}$$

which is the two-dimensional normal density with parameters $\left(-1,\ 3:\begin{pmatrix}1 & 0\\ 0 & 5\end{pmatrix}\right)$.

Another question for you (think really good and give your answer): the random variables X and Y have normal distributions with parameters, correspondingly, $(-1, 1)$ and $(3, 5)$; and these random variables are *uncorrelated.* What is the joint distribution of X, Y?

Somewhere above something was said that in the case of normal distribution, independence follows from being uncorrelated (formula (6)). So we are tempted to say that this will be the two-dimensional normal distribution with density (7).

But this would be wrong: again not necessarily. Formula (6) is again only true *under the assumption that the joint distribution is normal.* And *uncorrelatedness* is not about the joint distribution: it's just about some moments.

Counterexamples can be produced for any of the statements that I classified as "not necessarily true".

It turns out that a much more convenient way to handle multidimensional (and one-dimensional) normal distributions is using *characteristic functions*; or, which turned to be more popular with our students, using *moment-generating functions*.

If X is a (one-dimensional) random variable, its moment-generating function is defined by

$$M(t) = M_X(t) = E(e^{tX}). \tag{8}$$

The moment-generating function may be not defined for some real values of the argument t (and even for all, except $t = 0$, for which always $M_X(0) = 1$). Why these functions are useful, it is because of the following

Uniqueness Theorem. *If $M_X(t)$ and $M_Y(t)$ are defined at least in some interval of real line of nonzero length, and $M_X(t) = M_Y(t)$ in some interval of nonzero length, then the distributions of X and Y are the same: $\mu_X = \mu_Y$.*

The moment-generating function of the normal distribution with parameters (a, b) (which means: of a random variable having such a distribution) is

$$M(t) = e^{at+bt^2/2}. \tag{9}$$

A more convenient way to handle distributions is by means of *characteristic functions*. If X is a one-dimensional random variable, its characteristic function is defined by

$$\phi(z) = \phi_X(z) = E(e^{izX}). \tag{10}$$

In contrast to moment-generating functions, a characteristic function is defined at least for all real values of its argument z. This is because the random variable whose expectation we are considering is bounded:

$$|e^{izX}| = 1 \tag{11}$$

It is clear that the characteristic function is just the values of the moment-generating function on the imaginary axis, at $t = iz$.

The characteristic function of the normal distribution with parameters (a, b) is

$$\phi(z) = e^{iaz-bz^2/2} \tag{12}$$

(this, for z on the real axis, is not so easy to deduce from the fact that formula (9) holds for real values of t).

Uniqueness Theorem. *If $\phi_X(z)$ and $\phi_Y(z)$ coincide on the whole real axis, then the distributions of X and Y are the same: $\mu_X = \mu_Y$.*

To handle multidimensional distributions, we can introduce multidimensional moment-generating functions and multidimensional characteristic functions.

The n-dimensional moment-generating function $M_{\boldsymbol{X}}(\boldsymbol{t}) = M_{X_1,\,...,\,X_n}(t_1, ..., t_n)$ (the moment-generating function of a random vector, or the joint moment-generating function of the random variables X_1, ..., X_n) is defined as

$$M(\boldsymbol{t}) = M_{\boldsymbol{X}}(\boldsymbol{t}) = E(e^{\boldsymbol{t}\boldsymbol{\cdot}\boldsymbol{X}}), \tag{13}$$

where $\boldsymbol{t}\boldsymbol{\cdot}\boldsymbol{X}$ (with a boldfaced dot) is the dot product of two vectors. It can also be written as

$$\boldsymbol{t}\boldsymbol{\cdot}\boldsymbol{X} = \boldsymbol{t}^T\boldsymbol{X} \tag{14}$$

if we write $\boldsymbol{t}$ and $\boldsymbol{X}$ as *column* vectors, and apply matrix multiplication, the superscript ^{T} meaning "transposed". In components, the definition (13) can be written as

$$M(t_1, ..., t_n) = M_{X_1,\,...,\,X_n}(t_1, ..., t_n) = E\big(\exp\{\sum_{k=1}^{n} t_k \cdot X_k\}\big). \tag{15}$$

The n-dimensional characteristic function is defined as

$$\begin{aligned} \phi(\boldsymbol{z}) &= \phi_{\boldsymbol{X}}(\boldsymbol{z}) = E(\exp\{i\,\boldsymbol{z}\boldsymbol{\cdot}\boldsymbol{X}\}) = E(\exp\{i\,\boldsymbol{z}^T\boldsymbol{X}\}) = \phi(z_1, ..., z_n) \\ &= \phi_{X_1,\,...,\,X_n}(z_1, ..., z_n) = E\big(\exp\{\sum_{k=1}^{n} i\,z_k X_k\}\big). \end{aligned} \tag{16}$$

Uniqueness Theorem. *If $M_{\boldsymbol{X}}(\boldsymbol{t})$ and $M_{\boldsymbol{Y}}(\boldsymbol{t})$ are defined at least in some nonempty open subset of $\mathbb{R}^n$, and $M_{\boldsymbol{X}}(\boldsymbol{t}) = M_{\boldsymbol{Y}}(\boldsymbol{t})$ nonempty open set, then the distributions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the same: $\mu_{\boldsymbol{X}} = \mu_{\boldsymbol{Y}}$.*

Uniqueness Theorem. *If $\phi_{\boldsymbol{X}}(\boldsymbol{z})$ and $\phi_{\boldsymbol{Y}}(\boldsymbol{z})$ coincide on the whole space $\mathbb{R}^n$, then the distributions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the same: $\mu_{\boldsymbol{X}} = \mu_{\boldsymbol{Y}}$.*

It turns out that it is more convenient to handle multidimensional normal (Gaussian) distributions using moment-generating or characteristic functions.

Definition. Let $\boldsymbol{a}$ be an n-dimensional vector, B a symmetric nonnegative-definite $n \times n$ matrix. We say that a random vector $\boldsymbol{X}$ has a normal distribution with parameters $(\boldsymbol{a}, B)$ if its moment-generating function is given by

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \exp\{\boldsymbol{a}\boldsymbol{\cdot}\boldsymbol{t}+\boldsymbol{t}\boldsymbol{\cdot}B\boldsymbol{t}/2\} = \exp\{\boldsymbol{a}^T\boldsymbol{t}+\boldsymbol{t}^TB\boldsymbol{t}/2\} = \exp\{\sum_{k=1}^{n} a_k t_k + \frac{1}{2}\sum_{k,\,l=1}^{n} b_{kl}\,t_k t_l\}; \tag{17}$$

or, which is the same, if its characteristic function is given by

$$\phi_{\boldsymbol{X}}(\boldsymbol{z}) = \exp\{i\,\boldsymbol{a}\boldsymbol{\cdot}\boldsymbol{z}-\boldsymbol{z}\boldsymbol{\cdot}B\boldsymbol{z}/2\} = \exp\{i\,\boldsymbol{a}^T\boldsymbol{z}-\boldsymbol{z}^TB\boldsymbol{z}/2\} = \exp\{i\sum_{k=1}^{n} a_k z_k - \frac{1}{2}\sum_{k,\,l=1}^{n} b_{kl}\,z_k z_l\}. \tag{18}$$

Note that, while according to the definition (2), (3) an n-dimensional normal distribution is defined only in the case of the matrix B (the covariance matrix) being *positive*-definite, here we allow *nonnegative*-definite matrices. In the one-dimensional case the difference is negligible, because nonnegative numbers are just the positive numbers, and 0; but in the multidimensional case we have a whole spectrum of matrices with ranks from 0 to $n-1$.

It turns out that in the case of a *positive*-definite matrix B the new definition does not contradict the old one. Indeed, let $\boldsymbol{X}$ have the density (2). We have:

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \int\limits_{\mathbb{R}^n} \cdots \int \exp\{\boldsymbol{t}^T \boldsymbol{x}\} \cdot \frac{1}{(2\pi)^{n/2}\sqrt{\det(B)}} \exp\{-(\boldsymbol{x}-\boldsymbol{a})^T B^{-1} (\boldsymbol{x}-\boldsymbol{a})/2\}\, d\boldsymbol{x}. \tag{19}$$

Every positive-definite symmetric matrix can be represented as

$$B = AA^T, \tag{20}$$

where A is a nonsingular square matrix. We have: $Q = B^{-1} = (A^T)^{-1}A^{-1}$. Let us make in this integral the substitution

$$\boldsymbol{y} = A^{-1}(\boldsymbol{x}-\boldsymbol{a}), \qquad \boldsymbol{x} = \boldsymbol{a} + A\boldsymbol{y}. \tag{21}$$

The range of integration for $\boldsymbol{y}$ is still the whole $\mathbb{R}^n$; the Jacobian $\frac{d\boldsymbol{x}}{d\boldsymbol{y}}$ is equal to $\det(A) = \sqrt{\det(B)}$, so

$$\begin{aligned} M_{\boldsymbol{X}}(\boldsymbol{t}) &= \int\limits_{\mathbb{R}^n} \cdots \int \exp\{\boldsymbol{t}^T(\boldsymbol{a}+A\boldsymbol{y})\} \cdot \frac{1}{(2\pi)^{n/2}} \exp\{-(A\boldsymbol{y})^T B^{-1} A\boldsymbol{y}/2\}\, d\boldsymbol{y} \\ &= e^{\boldsymbol{a}^T\boldsymbol{t}} \int\limits_{\mathbb{R}^n} \cdots \int \frac{1}{(2\pi)^{n/2}} \exp\{\boldsymbol{t}^T A\boldsymbol{y} - (\boldsymbol{y}^T A^T (A^T)^{-1} A^{-1} A\boldsymbol{y})/2\}\, d\boldsymbol{y}. \end{aligned} \tag{22}$$

The matrix $A^T(A^T)^{-1}A^{-1}A$ is nothing but the identity matrix I, and

$$\begin{aligned} M_{\boldsymbol{X}}(\boldsymbol{t}) &= e^{\boldsymbol{a}^T\boldsymbol{t}} \int\limits_{\mathbb{R}^n} \cdots \int \frac{1}{(2\pi)^{n/2}} \exp\{\boldsymbol{t}^T A\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{y}/2\}\, d\boldsymbol{y} \\ &= e^{\boldsymbol{a}^T\boldsymbol{t} + |A^T\boldsymbol{t}|^2/2} \int\limits_{\mathbb{R}^n} \cdots \int \frac{1}{(2\pi)^{n/2}} \exp\{-|\boldsymbol{y} - A^T\boldsymbol{t}|^2/2\}\, d\boldsymbol{y}. \end{aligned} \tag{23}$$

Substitution $\boldsymbol{z} = \boldsymbol{y} - A^T\boldsymbol{t}$ with Jacobian equal to 1 leads to

$$\int\limits_{\mathbb{R}^n} \cdots \int \frac{1}{(2\pi)^{n/2}} \exp\{-|\boldsymbol{y} - A^T\boldsymbol{t}|^2/2\}\, d\boldsymbol{y} = \int\limits_{\mathbb{R}^n} \cdots \int \frac{1}{(2\pi)^{n/2}} e^{-|\boldsymbol{z}|^2/2}\, d\boldsymbol{z} = 1, \tag{24}$$

and

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = e^{\boldsymbol{a}^T\boldsymbol{t} + |A^T\boldsymbol{t}|^2/2} = e^{\boldsymbol{a}^T\boldsymbol{t} + \boldsymbol{t}^T AA^T\boldsymbol{t}/2} = e^{\boldsymbol{a}^T\boldsymbol{t} + \boldsymbol{t}^T B\boldsymbol{t}/2}, \tag{25}$$

which is the expression (17).

In the case that the covariance matrix B is singular, the normal distribution has no n-dimensional density

Of course, the question of *existence* pops out, especially in the case of singular nonnegative definite matrices. But before that, let us have the following

Microtheorem 1. *If an n-dimensional random vector $\boldsymbol{X}$ has a normal distribution with parameters $(\boldsymbol{a}, B)$; A is an $m \times n$ matrix, $\boldsymbol{c} \in \mathbb{R}^m$, and $\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{c}$, then the random vector $\boldsymbol{Y}$ also has a normal distribution.*

Proof. We have:

$$\begin{aligned} M_{\boldsymbol{Y}}(\boldsymbol{t}) &= E(e^{\boldsymbol{t}^T(A\boldsymbol{X}+\boldsymbol{c})}) = e^{\boldsymbol{c}^T\boldsymbol{t}} \cdot E(e^{(A^T\boldsymbol{t})^T\boldsymbol{X}}) = e^{\boldsymbol{c}^T\boldsymbol{t}} \cdot M_{\boldsymbol{X}}(A^T\boldsymbol{t}) \\ &= e^{\boldsymbol{c}^T\boldsymbol{t}} \cdot e^{\boldsymbol{a}^T(A^T\boldsymbol{t})+(A^T\boldsymbol{t})^T B(A^T\boldsymbol{t})/2} = e^{\boldsymbol{c}^T\boldsymbol{t}} \cdot e^{(A\boldsymbol{a})^T\boldsymbol{t}+\boldsymbol{t}^T(ABA^T)\boldsymbol{t}/2}. \end{aligned} \tag{26}$$

This is the moment-generating function of the normal distribution with parameters $(A\boldsymbol{a} + c,\ ABA^T)$ (note that ABA^T is a symmetric $m \times m$ matrix, as it should be).

Of course, we could have evaluated the *parameters* as the expectation and the covariance matrix: for the latter:

$$\begin{aligned} &E\big(\big(Y - E(Y)\big) \cdot \big(Y - E(Y)\big)^T\big) = E\big(A(X - \boldsymbol{a}) \cdot \big(A(X - \boldsymbol{a})\big)^T\big) \\ &\quad = E\big(A \cdot \big((X - \boldsymbol{a})(X - \boldsymbol{a})^T\big) \cdot A^T\big) = A \cdot E\big((X - \boldsymbol{a})(X - \boldsymbol{a})^T\big)\big) \cdot A^T = ABA^T. \end{aligned} \tag{27}$$

Microtheorem 2. *For every vector $\mathbf{a} \in \mathbb{R}^n$ and every symmetric nonnegative-definite $n \times n$ matrix B there exists a random vector having the normal distribution with parameters $(\boldsymbol{a}, B)$.*

Proof. There certainly exists a random vector $\boldsymbol{X}$ having the normal distribution with parameters $(\mathbf{0}, I)$: this is a vector with independent components having the standard normal distribution.

Every symmetric nonnegative-definite matrix can be diagonalized:

$$B = C \begin{pmatrix} d_{11} & 0 & ... & 0 \\ 0 & d_{22} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & d_{nn} \end{pmatrix} C^T. \tag{28}$$

Since the matrix is nonnegative-definite, the diagonal entries are nonnegative; we can take

$$A = C \begin{pmatrix} \sqrt{d_{11}} & 0 & ... & 0 \\ 0 & \sqrt{d_{22}} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \sqrt{d_{nn}} \end{pmatrix}, \tag{29}$$

and then

$$B = AA^T. \tag{30}$$

The random vector

$$\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{a} \tag{31}$$

has exactly the distribution we were talking about.

Microtheorem 3. *If a sequence $\boldsymbol{X}_n$ of r-dimensional random vectors with normal distributions with parameters $(\boldsymbol{a}_n, B_n)$ converges in probability to a random vector $\boldsymbol{X}$, the latter also has a normal distribution – with parameters* $(\lim_{n\to\infty} \boldsymbol{a}_n,\ \lim_{n\to\infty} B_n)$.

Proof. Moment-generating functions are not well adapted for this proof, so I am going to use characteristic functions. We have, for every $\boldsymbol{z} \in \mathbb{R}^r$:

$$e^{i\boldsymbol{z}^T\boldsymbol{X}} = \lim_{n\to\infty} (P)\, e^{i\boldsymbol{z}^T\boldsymbol{X}_n}. \tag{32}$$

All these random variables are dominated in absolute value by 1, and $E(1) = 1 < \infty$; so by the Dominated-Convergence Theorem (Theorem 2.2) we have:

$$\phi_{\boldsymbol{X}}(\boldsymbol{z}) = \lim_{n\to\infty} E(e^{i\boldsymbol{z}^T\boldsymbol{X}_n}) = e^{i(\lim_{n\to\infty}\boldsymbol{a}_n)^T\boldsymbol{z} - \boldsymbol{z}^T\cdot\lim_{n\to\infty}B_n\cdot\boldsymbol{z}/2}, \tag{33}$$

and this is the characteristic function of the normal distribution.

I have discovered that in one of the problems (Problem **15**) I used the words "converges weakly" that are not understandable for the students (I did not expect it: this is a standard part of a probability course). Let me speak about this.

We say that the sequence of distributions μ_{X_n} of random variables X_n converges weakly to the distribution μ_Y of a random variable Y (notation: $\mu_{X_n} \to_w \mu_Y$) if for every bounded continuous function $g(x)$

$$\lim_{n\to\infty} E\big(g(X_n)\big) = E\big(g(Y)\big). \tag{34}$$

It's somewhat awkward that this definition is formulated not in terms of distributions μ_{X_n}, μ_Y, but rather in those of the corresponding random variables. But this is because we have no unified way to treat all distributions: we treat discrete distributions one way, and continuous ones another way. For continuous distributions with densities p_{X_n}, p_Y (and this seems to be the case in Problem **15**) (34) can be rewritten as

$$\lim_{n\to\infty} \int g(x)\cdot p_{X_n}(x)\ dx = \int g(x)\cdot p_Y(x)\ dx \tag{35}$$

(the integrals are taken from $-\infty$ to ∞ in the case of number-valued random variables, or over $\mathbb{R}^r$ for r-dimensional random vectors).

If $p_{X_n}(x)$, $p_Y(x)$ are probability densities, and for every x

$$\lim_{n\to\infty} p_{X_n}(x) = p_Y(x), \tag{36}$$

then weak convergence (35) takes place. However weak convergence is much weaker than the convergence (36): it may be that (35) is satisfied, while (36) is not. Example:

$$p_{X_n}(x) = I_{[0,\,1]}(x)\cdot(1+\cos 2^n\pi x), \qquad p_Y(x) = I_{[0,\,1]}(x). \tag{37}$$

Theorem (usually taught in the elementary probability course). *For* $\mu_{X_n} \to_w \mu_Y$ *it is necessary and sufficient that the corresponding cumulative distribution functions*

$$F_{X_n}(t) = P\{X_n \le t\} = \mu_{X_n}(-\infty,\, t] \to F_Y(t) = P\{Y \le t\} = \mu_Y(-\infty,\, t] \tag{38}$$

at every point $t \in (-\infty, \infty)$ *at which the function* F_Y *is continuous.*

This is the way in which we can handle the example (37).

However, the easiest way to handle weak convergence is through characteristic functions:

Theorem (usually not taught in the elementary probability course, but should form a part of an advanced course). *For* $\mu_{X_n} \to_w \mu_Y$ *it is necessary and sufficient that the corresponding characteristic functions converge at every point*:

$$\lim_{n\to\infty} \phi_{X_n}(z) = \phi_Y(z). \tag{39}$$

This is true for one-dimensional random variables, and for multi-dimensional too.

Now the time seems to have come for solutions of concrete problems.

$\boxed{1}$ The random vector $(Z_{t_1},\, Z_{t_2})$ is a linear function of the random vector (X_1, Y_1, X_2, Y_2), which has a four-dimensional normal distribution (with parameters $(\mathbf{0}, I)$); so by Microtheorem 1 it also has a normal distribution – a two-dimensional one. Its first parameter, the mean value is, certainly, $\mathbf{0}$; and the covariance matrix

$$B_{t_1,\,t_2} = E\Big(\begin{pmatrix} Z_{t_1} \\ Z_{t_2} \end{pmatrix} \cdot \begin{pmatrix} Z_{t_1} & Z_{t_2} \end{pmatrix}\Big) = \begin{pmatrix} E(Z_{t_1}^2) & E(Z_{t_1} \cdot Z_{t_2}) \\ E(Z_{t_1} \cdot Z_{t_2}) & E(Z_{t_1}^2) \end{pmatrix}. \tag{40}$$

We have (using the fact that X_1, Y_1, X_2, Y_2 are independent):

$$E(Z_{t_1}^2) = E(X_1^2) \cdot \cos^2 t_1 + E(Y_1^2) \cdot \sin^2 t_1 + E(X_2^2) \cdot \cos^2 2t_1 + E(Y_2^2) \cdot \sin^2 2t_1 = 2, \tag{41}$$

similarly the second diagonal entry. The off-diagonal entry:

$$\begin{aligned} E(Z_{t_1} \cdot Z_{t_2}) = E(X_1^2) \cdot \cos t_1 \cos t_2 + E(Y_1^2) \cdot \sin t_1 \sin t_2 + E(X_2^2) \cdot \cos 2t_1 \cos 2t_2 \\ + E(Y_1^2) \cdot \sin 2t_1 \sin 2t_2 = \cos(t_2 - t_1) + \cos 2(t_2 - t_1). \end{aligned} \tag{42}$$

The matrix

$$B_{t_1,\,t_2} = \begin{pmatrix} 2 & \cos(t_2 - t_1) + \cos 2(t_2 - t_1) \\ \cos(t_2 - t_1) + \cos 2(t_2 - t_1) & 2 \end{pmatrix} \tag{43}$$

is nonsingular if $t_2 - t_1 \ne 2\pi k$, and we can write the density:

$$\begin{aligned} \det(B) &= 4 - \big(\cos(t_2 - t_1) + \cos 2(t_2 - t_1)\big)^2, \\ B^{-1} &= \frac{1}{\det(B)} \begin{pmatrix} 2 & -\big(\cos(t_2 - t_1) + \cos 2(t_2 - t_1)\big) \\ -\big(\cos(t_2 - t_1) + \cos 2(t_2 - t_1)\big) & 2 \end{pmatrix}, \end{aligned} \tag{44}$$

$$p_{Z_{t_1},\,Z_{t_2}}(z_1,z_2) = \frac{1}{2\pi\sqrt{4-\bigl(\cos(t_2-t_1)+\cos 2(t_2-t_1)\bigr)^2}}\times$$
$$\times\exp\Bigl\{-\frac{2z_1^2-2\bigl(\cos(t_2-t_1)+\cos 2(t_2-t_1)\bigr)z_1z_2+2z_2^2}{2\bigl(4-\bigl(\cos(t_2-t_1)+\cos 2(t_2-t_1)\bigr)^2\bigr)}\Bigr\}. \tag{45}$$

This could be obtained without using any moment-generating or characteristic functions, just passing from a four-dimensional density to a two-dimensional one (but would require honest and pretty heavy calculations).

For $t_2 - t_1 = 2\pi k$ we have, of course, $Z_{t_2} = Z_{t_1}$, and the distribution of this random variable is normal with parameters $(0, 2)$ (the joint distribution is concentrated on the line $z_2 = z_1$).

One-, and two-, and three-, and four-dimensional joint distributions of the random variables Z_t can have densities, but not five-dimensional: because the covariance matrix of these random variables is equal to $A\,B_{X_1,\,Y_1,\,X_2,\,Y_2}A^T$, and cannot have rank greater than 4.

For the concrete values $t_1 = \pi/6$, $t_2 = \pi/2$ we have:

$$\cos(t_2-t_1)+\cos 2(t_2-t_1) = \cos\pi/3 - \cos 2\pi/3 = 1/2 - 1/2 = 0; \tag{46}$$

so the random variables $Z_{\pi/6}$, $Z_{\pi/2}$ happen to be independent and have the joint density

$$p_{Z_{\pi/6},\,Z_{\pi/2}}(z_1,z_2) = \frac{1}{4\pi}\,e^{-(z_1^2+z_2^2)/4}. \tag{47}$$

But if you wrote that $Z_{\pi/6}$, $Z_{\pi/2}$ have joint normal density without any reason, or by the reason that each of them has normal distribution, this is wrong: *not* for that reason; or if you wrote that $Z_{\pi/6}$, $Z_{\pi/2}$ are independent without giving any reason, or by the reason that *every* two random variables are independent unless we are told otherwise, this is also wrong: *not* for that reason.

2 We have for nonnegative integer y_1, y_2:

$$P\{X_1 - X_0 = y_1,\, X_3 - X_1 = y_2\} = P\{X_1 = y_1,\, X_3 = y_1 + y_2\} = \pi_\lambda(y_1)\cdot\pi_{2\lambda}(y_2). \tag{48}$$

This means that the random variables $X_1 - X_0$, $X_3 - X_1$ are independent and have Poissonian distributions with parameters, respectively, λ and 2λ.

3 We have to check that conditions $2_1)-2_n)$ of Lecture note 3 (formulas (3.1)) are satisfied. These conditions, for the discrete case, are written as:

$$\begin{aligned}
&\sum_{x_1\in\mathrm{Sp}} p_{t_1,\,t_2,\,\dots,\,t_n}(x_1,x_2,\dots,x_n) = p_{t_2,\,\dots,\,t_n}(x_2,\dots,x_n),\\
&\dots\dots\dots\dots\\
&\sum_{x_i\in\mathrm{Sp}} p_{t_1,\,\dots,\,t_{i-1},\,t_i,\,t_{i+1},\,\dots,\,t_n}(x_1,\dots,x_{i-1},x_i,x_{i+1},\dots,x_n)\\
&\qquad\qquad = p_{t_1,\,\dots,\,t_{i-1},\,t_{i+1},\,\dots,\,t_n}(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n),\\
&\dots\dots\dots\dots\\
&\sum_{x_i\in\mathrm{Sp}} p_{t_1,\,\dots,\,t_{n-1},\,t_n}(x_1,\dots,x_{n-1},x_n) = p_{t_1,\,\dots,\,t_{n-1}}(x_1,\dots,x_{n-1})
\end{aligned} \tag{49}$$

(the discrete analogue of formulas (3.15)).

The last of them is satisfied for the distributions (3_1) because $\sum_y \pi_{\lambda(t_n - t_{n-1})}(y) = 1$; the i-th, $i = 1, ..., n-1$, because

$$\begin{aligned}
&\sum_y \pi_{\lambda(t_i - t_{i-1})}(y) \cdot \pi_{\lambda(t_{i+1} - t_i)}(x - y) \\
&\quad = \sum_{y=0}^{x} \frac{\big(\lambda(t_i - t_{i-1})\big)^y e^{-\lambda(t_i - t_{i-1})}}{y!} \cdot \frac{\big(\lambda(t_{i+1} - t_i)\big)^{x-y} e^{-\lambda(t_{i+1} - t_i)}}{(x-y)!} \\
&\quad = \frac{\lambda^x e^{-\lambda(t_{i+1} - t_{i-1})}}{x!} \sum_{y=0}^{x} \binom{x}{y} (t_i - t_{i-1})^x (t_{i+1} - t_i)^{x-y} \\
&\quad = \frac{\big(\lambda(t_{i+1} - t_{i-1})\big)^x e^{-\lambda(t_{i+1} - t_{i-1})}}{x!}
\end{aligned} \tag{50}$$

(the distribution of the sum of two independent Poisson random variables with parameters $\lambda(t_i - t_{i-1})$ and $\lambda(t_{i+1} - t_i)$ is Poisson with parameter $\lambda(t_{i+1} - t_{i-1})$).

So the consistency conditions are satisfied, and there exists a stochastic process with finite-dimensional distributions (3_1).

This process is called *the Poisson process with parameter* λ.

As for whether it can have continuous trajectories: of course not: no non-constant function taking integer values can be continuous: it can change only with jumps.

4 Denote $Y_1 = X_1 = X_1 - X_0$, $Y_2 = X_3 - X_1$. We have: $\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = A\boldsymbol{X} = A \cdot \begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$, where $A = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$; this is a one-to-one transformation of $\mathbb{R}^2$. For $C \subseteq \mathbb{R}^2$

$$P\{\boldsymbol{Y} \in C\} = P\{\boldsymbol{X} \in A^{-1}C\} = \iint_{A^{-1}C} p_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}. \tag{51}$$

Making the substitution $\boldsymbol{y} = A\boldsymbol{x}$, $\boldsymbol{x} = A^{-1}\boldsymbol{y}$, $d\boldsymbol{x} = |\det(A^{-1})|\, d\boldsymbol{y}$ ($A^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$), we get:

$$P\{\boldsymbol{Y} \in C\} = \iint_{C} p_{\boldsymbol{X}}(A^{-1}\boldsymbol{y})\, d\boldsymbol{y}. \tag{52}$$

This means that the joint density of the random variables Y_1, Y_2 is

$$p_{Y_1, Y_2}(y_1, y_2) = p_{\boldsymbol{Y}}(\boldsymbol{y}) = p_{\boldsymbol{X}}(A^{-1}\boldsymbol{y}) = p_{X_1, X_3}(y_1, y_1 + y_2) = q_1(y_1) \cdot q_2(y_2). \tag{53}$$

That is, the random variables $X_1 - X_0$, $X_3 - X_1$ are independent and have, respectively, the uniform distribution on the interval $[0, 1]$, and uniform distribution on the interval $[0, 2]$.

5 Since the distribution of the sum of two independent random variables having uniform distributions on the intervals $[0, 1]$, $[0, 2]$ is not the uniform distribution on the interval $[0, 3]$, but rather has the density

$$p(x) = \begin{cases} x/2, & 0 \le x \le 1, \\ 1/2, & 1 \le x \le 2, \\ (3-x)/2, & 2 \le x \le 3, \\ 0, & x \notin [0,3], \end{cases} \tag{54}$$

we have:

$$\int_{-\infty}^{\infty} q_1(x_1) \cdot q_2(x_3 - x_1)\, dx_1 \ne q_3(x_3), \tag{55}$$

$$\mu_{1,3}(\mathbb{R}^1 \times C) = \int_C \Big[\int_{-\infty}^{\infty} q_1(x_1) \cdot q_2(x_3 - x_1)\, dx_1 \Big]\, dx_3 \ne \int_C q_3(x_3)\, dx_3 = \mu_3(C); \tag{56}$$

the consistency condition is not satisfied, there is *no* such stochastic process.

Of course, the question about whether it has continuous trajectories does not arise.

6 I have spoken about this in the lecture. The mean-square limit exists and is equal to 0. See also Problem **8**.

7 For $0 \le t < s$ we have:

$$E\big((f(s,\omega) - f(t,\omega))^2\big) = P\{t < T \le s\} = \int_t^s p_T(u)\, du \le s - t. \tag{57}$$

For arbitrary $t,\ s \in [0, \infty)$

$$E\big((f(s,\omega) - f(t,\omega))^2\big) \le |s - t| \to 0 \quad (s \to t). \tag{58}$$

So the random function $f(t,\omega)$ *is* mean-square continuous.

Also it is continuous in probability, because mean-square convergence implies convergence in probability.

8 Let us consider the finite-dimensional distributions of the process $\tilde{\boldsymbol{W}}_t$: $\mu_{\tilde{\boldsymbol{W}}_{t_1}, \tilde{\boldsymbol{W}}_{t_2}, \dots, \tilde{\boldsymbol{W}}_{t_n}}$. The $(n \cdot r)$-dimensional random vector $\tilde{\mathbb{W}} = \begin{pmatrix} \boldsymbol{W}_{t_1} \\ \dots \\ \boldsymbol{W}_{t_n} \end{pmatrix}$ is expressed as the matrix product:

$$\tilde{\mathbb{W}} = \mathbb{O} \cdot \mathbb{W}, \tag{59}$$

where $\mathbb{W} = (\,\boldsymbol{W}_{t_1}, \dots \boldsymbol{W}_{t_n}\,)$, and the $(n\,r \times n\,r)$ matrix $\mathbb{O}$ is given by

$$\mathbb{O} = \begin{pmatrix} O & 0 & \dots & 0 \\ 0 & O & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & O \end{pmatrix}. \tag{60}$$

The mean value of the random vector $\mathbb{W}$ is $\begin{pmatrix} \boldsymbol{x}_0 \\ ... \\ \boldsymbol{x}_0 \end{pmatrix}$, and its covariance matrix is

$$\mathbb{B} = \begin{pmatrix} (t_1 - t_0)I & 0 & ... & 0 \\ 0 & (t_2 - t_1)I & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & (t_n - t_{n-1})I \end{pmatrix}. \tag{61}$$

By Microtheorem 1 above, the random vector $\tilde{\mathbb{W}}$ has the normal distribution with parameters $\mathbb{O} \cdot E(\mathbb{W}) = \begin{pmatrix} O\boldsymbol{x}_0 \\ ... \\ O\boldsymbol{x}_0 \end{pmatrix}$ (which is not surprising since the process $\tilde{W}_t$ starts from the point $O\boldsymbol{x}_0$) and

$$\begin{aligned} \mathbb{O} \cdot \mathbb{B} \cdot \mathbb{O}^T &= \begin{pmatrix} O \cdot (t_1 - t_0)I \cdot O^T & 0 & ... & 0 \\ 0 & O \cdot (t_2 - t_1)I \cdot O^T & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & O \cdot (t_n - t_{n-1})I \cdot O^T \end{pmatrix} \\ &= \begin{pmatrix} (t_1 - t_0)I & 0 & ... & 0 \\ 0 & (t_2 - t_1)I & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & (t_n - t_{n-1})I \end{pmatrix} = \mathbb{B}. \end{aligned} \tag{62}$$

So the finite-dimensional distributions are such as they should be for the Wiener process starting from the point $\boldsymbol{x}_0$ at time $t = 0$. What else do we need to state that $\tilde{\boldsymbol{W}}_t$ is a Wiener process? Ah, yes, continuity of the trajectories; but this is certainly there.

Returning to Problem **6**: if W_t^1, W_t^2 are two components of a two-dimensional Wiener process, take the orthogonal matrix $O = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$; we see that $\tilde{W}_t^1 = (W_t^1 + W_t^2)/\sqrt{2}$, $\tilde{W}_t^2 = (W_t^1 - W_t^2)/\sqrt{2}$ are two components of another two-dimensional Wiener process, and as such they are both one-dimensional Wiener processes. We have:

$$\sum_{i=1}^{n} (W_{t_i}^1 - W_{t_{i-1}}^1)(W_{t_i}^2 - W_{t_{i-1}}^2) = \frac{1}{2} \sum_{i=1}^{n} [(\tilde{W}_{t_i}^1 - \tilde{W}_{t_{i-1}}^1)^2 - (\tilde{W}_{t_i}^2 - \tilde{W}_{t_{i-1}}^2)^2] \tag{63}$$

and this has mean-squares limit equal to $\frac{1}{2}[(b-a) - (b-a)] = 0$.

9 Take $X_t = v_0 + \sigma \int_{t_0}^{t} e^{a(s-t_0)}\, dW_s$, $dX_t = \sigma\, e^{a(t-t_0)}\, dW_t$, $F(t,x) = e^{-a(t-t_0)} \cdot x$, then $V_t = F(t, X_t)$. We have: $\frac{\partial F}{\partial t}(t,x) = -a \cdot e^{-a(t-t_0)} \cdot x$, $\frac{\partial F}{\partial x}(t,x) = e^{-a(t-t_0)}$, $\frac{\partial^2 F}{\partial x^2}(t,x) = 0$. By Itô's formula,

$$dV_t = -a \cdot e^{-a(t-t_0)} \cdot X_t\, dt + e^{-a(t-t_0)} \cdot \sigma\, e^{a(t-t_0)}\, dW_t = -a \cdot V_t\, dt + \sigma\, dW_t. \tag{64}$$

10 The stochastic integral is equal to

$$\sum_{i=1}^{n} f(t_{i-1}) \cdot (W_{t_i} - W_{t_{i-1}}). \tag{65}$$

This is a sum of n independent normal random variables, and as such this random variable has the normal distribution with parameters 0 and

$$\sum_{i=1}^{n} f(t_{i-1})^2 \cdot \mathrm{Var}(W_{t_i} - W_{t_{i-1}}) = \sum_{i=1}^{n} f(t_{i-1})^2 \cdot (t_i - t_{i-1}) = \int_a^b f(t)^2\, dt. \tag{66}$$

11 The parameters are $\big(0, \int_a^b f(t)^2\, dt\big)$.

12 We have: $\int_a^b X_t\, dt = \underset{\max_{1\le i\le n}(t_i - t_{i-1})\to 0}{\text{l.i.m.}} \sum_{i=1}^{n} X_{t_i^*} \cdot (t_i - t_{i-1})$. According to Microtheorem 1 above, the distribution of the random variable $\sum_{i=1}^{n} X_{t_i^*} \cdot (t_i - t_{i-1})$ is normal; its characteristic function is equal to $E^{iaz - bz^2/2}$ with some a and b: the exponential function of a quadratic function of z. The characteristic function of the limiting random variable, $\int_a^b X_t\, dt$ is equal to the limit of characteristic functions of $\sum_{i=1}^{n} X_{t_i^*} \cdot (t_i - t_{i-1})$, and is the exponential function of a linear function of z too: a characteristic function of some normal distribution.

13 The distribution of the random variable V_t, which is a constant plus a stochastic integral of a non-random function $g(t)$, is the (weak) limit of the same constant plus the sum $\sum_{i=1}^{n} g(t_{i-1}) \cdot (W_{t_i} - W_{t_{i-1}})$. The sum, being one of independent normal random variables, has a normal distribution with parameters $\big(0, \sum_{i=1}^{n} g(t_{i-1})^2 \cdot (t_i - t_{i-1})\big)$; so the random variable V_t also has a normal distribution – with parameters $\big(v_0 e^{-a(t-t_0)}, \int_{t_0}^{t} \sigma^2 \cdot e^{-2a(t-s)}\, ds\big) = \big(0, \sigma^2 \cdot \dfrac{1 - e^{-2a(t-t_0)}}{2a}\big)$.

13 a We have:

$$\begin{aligned} X_t = x_0 + \int_{t_0}^{t} V_s\, ds &= x_0 + \int_{t_0}^{t} \Big[v_0 e^{-a(s-t_0)} + \sigma \int_{t_0}^{s} e^{-a(s-u)}\, dW_u \Big]\, ds \\ &= x_0 + v_0 \cdot \frac{1 - e^{-a(t-t_0)}}{a} + \sigma \int_{t_0}^{t} \Big[\int_{t_0}^{s} e^{-a(s-u)}\, dW_u \Big]\, ds. \end{aligned} \tag{67}$$

The Riemann integral of the stochastic integral is equal to a stochastic integral of a Riemann integral:

$$\begin{aligned} \int_{t_0}^{t} \Big[\int_{t_0}^{s} e^{-a(s-u)}\, dW_u \Big]\, ds &= \int_{t_0}^{t} \Big[\int_{t_0}^{t} I_{\{t_0 \le u \le s \le t\}} \cdot e^{-a(s-u)}\, dW_u \Big]\, ds \\ &= \int_{t_0}^{t} \Big[\int_{t_0}^{t} I_{\{t_0 \le u \le s \le t\}} \cdot e^{-a(s-u)}\, ds \Big]\, dW_u \\ &= \int_{t_0}^{t} \Big[\int_{u}^{t} e^{-a(s-u)}\, ds \Big]\, dW_u = \int_{t_0}^{t} \frac{1 - e^{-a(t-u)}}{a}\, dW_u. \end{aligned} \tag{68}$$

Why can we change the order of stochastic and Riemann integration? We have:

$$\int_a^b \Big[\int_c^d f(s,u)\ ds\Big]\ dW_u = \underset{\max(u_i-u_{i-1})\to 0}{\text{l.i.m.}} \sum_{i=1}^n \Big[\int_c^d f(s,u_{i-1})\ ds\Big] \cdot (W_{u_i} - W_{u_{i-1}})$$
$$= \int_c^d \big[\underset{\max(u_i-u_{i-1})\to 0}{\text{l.i.m.}} \sum_{i=1}^n f(s,u_{i-1}) \cdot (W_{u_i} - W_{u_{i-1}})\big]\ ds = \int_c^d \Big[\int_c^d f(s,u)\ dW_u\Big]\ ds. \tag{69}$$

By (67), (68), and Problem **11**, the random variable X_t has a normal distribution with parameters $x_0 + v_0 \cdot \dfrac{1-e^{-a(t-t_0)}}{a}$ and

$$\sigma^2 \int_{t_0}^t \Big(\frac{1-e^{-a(t-u)}}{a}\Big)^2\ du = \sigma^2 \cdot \frac{(t-t_0) - 2e^{-a(t-t_0)}/a + e^{-2a(t-t_0)}/2a}{a^2}. \tag{70}$$

15 Sorry, it turns out that we should not take $a=\sigma^2$, but rather $a=\sigma$. And the limiting distribution of $X_t - x_0$ is normal with parameters $(0,\ t-t_0)$: the same as for the Wiener process.

It can be checked (but this requires many calculations) that all finite-dimensional distributions of the process X_t converge to the same of the Wiener process starting from the point x_0.

16 Make a picture of the hexagon first.

The joint density $p_{X,Y}(x,y)$ is given by

$$p_{X,Y}(x,y) = \begin{cases} 1/3 & \text{if } 0 \le y \le 1,\ -1 \le x \le 1-y \text{ or } -1 \le y \le 0,\ -1-y \le x \le 1, \\ 0 & \qquad \text{otherwise.} \end{cases} \tag{71}$$

The density of the random variable Y taken separately

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\ dx = \begin{cases} (2-y)/3, & 0 \le y \le 1, \\ (2+y)/3, & -1 \le y \le 0, \\ 0, & y \notin [-1,1] \end{cases} \tag{72}$$

(draw a graph of this density). The conditional density $p_{X|Y=y}(x)$ is given by

$$p_{X|Y=y}(x) = \begin{cases} 1/(2-y) & \text{for } 0 \le y \le 1, -1 \le x \le 1-y, \\ 1/(2+y) & \text{for } -1 \le y \le 0, -1-y \le x \le 1, \\ \text{not important what} & \text{for } y \notin [-1,1]. \end{cases} \tag{73}$$

The conditional expectation

$$E\{X|Y=y\} = \int_{-\infty}^{\infty} x \cdot p_{X|Y=y}(x)\, dx = \begin{cases} \dfrac{1}{2-y} \displaystyle\int_{-1}^{1-y} x\, dx, & 0 \le y \le 1, \\ \dfrac{1}{2+y} \displaystyle\int_{-1-y}^{1} x\, dx, & -1 \le y \le 0, \\ \text{whatever}, & y \notin [-1,1], \end{cases}$$
$$= \begin{cases} -y/2, & -1 \le y \le 1, \\ \text{whatever}, & y \notin [-1,1]. \end{cases} \tag{74}$$

So as one version of $E(X\|Y)$ we can take $-Y/2$.

17 For $0 \le t < s$ we have to find

$$P\{Y_s = 1|Y_u = y_u,\, 0 \le u \le t\}, \qquad P\{Y_s = 0|Y_u = y_u,\, 0 \le u \le t\}. \tag{75}$$

We have to find these conditional probabilities only for functions y_u, $0 \le u \le t$, of one of the following types:

$$y_u = \begin{cases} 1, & 0 \le u < t_*, \\ 0, & t_* \le u \le t, \end{cases} \tag{76}$$

where $t_* \in [0,t]$; or

$$y_u \equiv 1, \qquad 0 \le u \le t. \tag{77}$$

For the functions y_u of the first type (formula (76)), obviously,

$$P\{Y_s = 1|Y_u = y_u,\, 0 \le u \le t\} = 0, \qquad P\{Y_s = 1|Y_u = y_u,\, 0 \le u \le t\} = 1. \tag{78}$$

If a function of the second type is observed, this gives us only the information that $T > t$, and

$$P\{Y_s = 1|Y_u \equiv 1,\, u \le t\} = P\{Y_s = 1|T > t\} = P\{T > s|T > t\} = \frac{e^{-as}}{e^{at}} = e^{a(s-t)},$$
$$P\{Y_s = 0|Y_u \equiv 1,\, u \le t\} = 1 - e^{-a(s-t)}. \tag{79}$$

In both cases the conditional probability depends only on the value of y_t: the case of trajectories of the form (76) is described as $y_t = 0$, and (77) as $y_t = 1$. So the process is a Markov one. The transition matrix is

$$P^{ts} = \begin{pmatrix} 1 & 0 \\ 1 - e^{-a(s-t)} & e^{-a(s-t)} \end{pmatrix} \tag{80}$$

(the rows and the columns are numbered in the order 0, 1).

This Markov process is time-homogeneous: the matrix (80) depends only on the difference $s - t$.

18 The same as the previous problem, but

$$P^{ts} = \begin{pmatrix} 1 & 0 \\ 1 - \dfrac{\displaystyle\int_s^\infty p_T(u)\,du}{\displaystyle\int_t^\infty p_T(u)\,du} & \dfrac{\displaystyle\int_s^\infty p_T(u)\,du}{\displaystyle\int_t^\infty p_T(u)\,du} \end{pmatrix}. \tag{81}$$

This is a Markov process, but not a time-homogeneous one.

19 Checking that is pure algebra.

What could I say about the corresponding Markov process? It turns out that there exists its version with right-continuous trajectories $X_t(\omega)$ (of course, *not* with *continuous* trajectories!). If X_t starts at 1 at time $t = 0$, there is a sequence of independent random variables T_1, T_2, T_3, T_4, ... having exponential distributions with parameters 1, 2, 1, 2, ...; the process spends the time T_1 at the point 1, and jumps to the point 2, spends the time T_2 there, jumps to 1, spends the time T_3 at this point, etc. If $X_0 = 2$, the parameters of the exponential distributions are 2, 1, 2, 1, ...

20 We have for $t \geq 0$:

$$\{T \leq t\} = \{Y_t = 0\}; \tag{82}$$

yes. As for $\min(T, t_*)$, it is Example 26.2.

21 For $0 \leq t < s$ we have to evaluate $E\{R_s | Y_u = y_u,\, u \leq t\}$ for y_u, $u \leq t$, jumping down at some time $t_* \in [0, t]$, and also for $y_u \equiv 1$, $u \leq t$. For functions y_u, $u \leq t$, of the first type, we have $T \leq t$, $R_s = R_t = Y_T + aT = aT$, and

$$E\{R_s | Y_u = y_u,\, u \leq t\} = aT = R_t. \tag{83}$$

For $y_u \equiv 1$, $u \leq t$, we have: $R_t = 1 + at$,

$$E\{R_s | Y_u = 1,\, u \leq t\} = E\{R_s | T > t\}. \tag{84}$$

The random variable R_s is given by

$$R_s = \begin{cases} 1 + as, & T > s, \\ aT, & T \leq s. \end{cases} \tag{85}$$

The conditional expectation (84) is equal to

$$\begin{aligned} &\frac{(1+as)\cdot P\{T > s\} + \displaystyle\int_t^s au \cdot ae^{-au}\,du}{P\{T > t\}} \\ &= \frac{(1+as)e^{-as} + ate^{-at} - ase^{-as} + e^{-at} - e^{-as}}{e^{-at}} \\ &= 1 + at = R_t. \end{aligned} \tag{86}$$

22 We have:

$$
\begin{aligned}
E(R_{\min(T,\,t_*)}) &= E(I_{\{T\le t_*\}}\cdot R_{\min(T,\,t_*)}) + E(I_{\{T>t_*\}}\cdot R_{\min(T,\,t_*)})\\
&= E(I_{\{T\le t_*\}}\cdot R_T) + E(I_{\{T>t_*\}}\cdot R_{t_*})\\
&= E(I_{\{T\le t_*\}}\cdot aT) + E\big(I_{\{T>t_*\}}\cdot (1+at_*)\big)\\
&= \int_0^{t_*} at\cdot p_T(t)\;dt + (1+at_*)\cdot P\{T>t_*\}\\
&= 1-(1+at_*)e^{-at_*} + (1+at_*)e^{-at_*} = 1;
\end{aligned} \tag{87}
$$

$$
E(R_{t_*}) = E(I_{\{T\le t_*\}}\cdot R_{t_*}) + E(I_{\{T>t_*\}}\cdot R_{t_*}) = E(I_{\{T\le t_*\}}\cdot R_T) + E(I_{\{T>t_*\}}\cdot R_{t_*}), \tag{88}
$$

which is the same as (87), that is, equal to 1.

23 We have:

$$
P\{L_t \ne R_t\} = P\{T=t\} = \int_t^t p_T(u)\;du = 0. \tag{89}
$$

Since the expectations, or conditional expectations, do not change if we change the random variable only with zero probability, we have that almost surely

$$
E(L_s \| Y_u,\, u\le t) = E(R_s \| Y_u,\, u\le t) = R_t = L_t, \tag{90}
$$

the random function L_t *is* a martingale.

24 We have $E(L_{t_*}) = E(R_{t_*}) = 1$; and

$$
\begin{aligned}
E(L_{\min(T,\,t_*)}) &= E(I_{\{T\le t_*\}}\cdot L_{\min(T,\,t_*)}) + E(I_{\{T>t_*\}}\cdot L_{\min(T,\,t_*)})\\
&= E(I_{\{T\le t_*\}}\cdot L_T) + E(I_{\{T>t_*\}}\cdot L_{t_*})\\
&= E\big(I_{\{T\le t_*\}}\cdot (1+aT)\big) + E\big(I_{\{T>t_*\}}\cdot (1+at_*)\big)\\
&= \int_0^{t_*} (1+at)\cdot p_T(t)\;dt + (1+at_*)\cdot P\{T>t_*\}\\
&= 1 + (1-e^{-at_*}) > 1.
\end{aligned} \tag{91}
$$

25 For a finite interval $(a,\,c)$ we solve the boundary-value problem

$$
\begin{gathered}
\tfrac12\, u''(x) + bu'(x) = 0\\
u(a)=0,\;\; u(c)=1:
\end{gathered}
$$

$$
u(x) = \frac{1-e^{-2b(x-a)}}{1-e^{-2b(c-a)}}, \qquad a\le x\le c,
$$

finding $P\{X^x_{\tau^x_{(a,\,c)}} = c\}$;

$$
P\{X^x_{\tau^x_{(a,\,c)}} = a\} = 1 - P\{X^x_{\tau^x_{(a,\,c)}} = c\} = \frac{e^{2b(c-x)}-1}{e^{2b(c-a)}-1}.
$$

Taking $c \to \infty$, we get:

$$P\{\tau^x_{(a,\,\infty)} < \infty\} = \lim_{c\to\infty} P\{X^x_{\tau^x_{(a,\,c)}} = a\} = \lim_{c\to\infty} \frac{e^{2b(c-x)}-1}{e^{2b(c-a)}-1} = e^{-2b(x-a)};$$

taking $a \to -\infty$:

$$P\{\tau^x_{(-\infty,\,c)} < \infty\} = \lim_{a\to-\infty} P\{X^x_{\tau^x_{(a,\,c)}} = c\} = \lim_{a\to-\infty} \frac{1-e^{-2b(x-a)}}{1-e^{-2b(c-a)}} = 1.$$

This is quite natural: for the Wiener process the time $\tau^x_{(-\infty,\,c)}$ is finite almost surely, so the same will be for the process $W_t + bt$ that is to the right of W_t for every $t > 0$.

26 The boundary-value problem for $m(x) = E(\tau^x_{(a,\,c)})$ is

$$\tfrac{1}{2}\, m''(x) + bm'(x) = -1,$$
$$m(a) = m(c) = 0.$$

The general solution of the equation without the boundary conditions is

$$m(x) = -x/b + C_1 + C_2 e^{-2bx},$$

where C_1, C_2 are arbitrary constants. Taking boundary conditions into account, we get:

$$C_2 = -\frac{c-a}{b(e^{-2ba}-e^{-2bc})}, \qquad C_1 = \frac{ce^{-2ba}-ae^{-2bc}}{b(e^{-2ba}-e^{-2bc})},$$

$$E(\tau^x_{(a,\,c)}) = m(x) = -\frac{x}{b} + \frac{c(e^{-2ba}-e^{-2bx}) + a(e^{-2bx}-e^{-2bc})}{b(e^{-2ba}-e^{-2bc})}.$$

27 We have:

$$E(\tau^x_{(-\infty,\,c)}) = \lim_{a\to-\infty} E(\tau^x_{(a,\,c)}) = \lim_{a\to-\infty} \Big[-\frac{x}{b} + \frac{c(e^{-2ba}-e^{-2bx}) + a(e^{-2bx}-e^{-2bc})}{b(e^{-2ba}-e^{-2bc})}\Big]$$
$$= \lim_{a\to-\infty} \Big[-\frac{x}{b} + \frac{c(1-e^{-2b(x-a)}) + a(e^{-2b(x-a)}-e^{-2b(c-a)})}{b(1-e^{-2b(c-a)})}\Big] = \frac{c-x}{b}.$$

28 Just plain differentiation.

29 The expectation $E\big(u(\boldsymbol{W}^{\boldsymbol{x}}_t)\big)$ is equal to

$$\iint\limits_{\mathbb{R}^2} (-\ln|\boldsymbol{y}|)\cdot \frac{1}{2\pi t}\, e^{-|\boldsymbol{y}-\boldsymbol{x}|^2/2t}\, d\boldsymbol{y}.$$

This integral is an improper one in two ways: near the point $\boldsymbol{y} = \boldsymbol{0}$, and at the infinity.

The integral over a circle of radius ρ centered at $\mathbf{0}$ does not exceed

$$\frac{1}{2\pi t}\iint\limits_{\mathbb{R}^2}\big|\ln|\boldsymbol{y}|\big|\,d\boldsymbol{y}=\frac{1}{2\pi t}\int_0^{2\pi}\Big[\int_0^{\rho}|\ln r|\cdot r\,dr\Big]\,d\theta<\infty.$$

At infinity, the exponential function goes to 0 so fast, and the logarithmic function so slow that the integral converges at infinity.

So: the expectation is finite.

30 The two-dimensional normal distribution with the mean $\boldsymbol{x}$ and the covariance matrix $\begin{pmatrix} t & 0\\ 0 & t\end{pmatrix}$ becomes "wider and wider" as $t\to\infty$, going more and more into the parts of the plane where the function u takes large negative values; so it seems that $\lim_{t\to\infty}E\big(u(\boldsymbol{W}_t)\big)=\infty$. Let us prove this.

For a large C take a circle of radius e^{C+1} centered at $\mathbf{0}$; for $\boldsymbol{y}$ outside this circle $u(\boldsymbol{y})<-C-1$. We have:

$$\iint\limits_{C}u(\boldsymbol{y})\cdot\frac{1}{2\pi t}\,e^{-|\boldsymbol{y}-\boldsymbol{x}|^2/2t}\,d\boldsymbol{y}=\iint\limits_{\{\boldsymbol{y}\colon|\boldsymbol{y}|\ge e^{C+1}\}}u(\boldsymbol{y})\cdot\frac{1}{2\pi t}\,e^{-|\boldsymbol{y}-\boldsymbol{x}|^2/2t}\,d\boldsymbol{y}+\iint\limits_{\{\boldsymbol{y}\colon|\boldsymbol{y}|< e^{C+1}\}}u(\boldsymbol{y})\cdot\frac{1}{2\pi t}\,e^{-|\boldsymbol{y}-\boldsymbol{x}|^2/2t}\,d\boldsymbol{y}.$$

The first integral in the right-hand side is less than

$$(-C-1)\cdot\iint\limits_{\{\boldsymbol{y}\colon|\boldsymbol{y}|\ge e^{C+1}\}}\frac{1}{2\pi t}\,e^{-|\boldsymbol{y}-\boldsymbol{x}|^2/2t}\,d\boldsymbol{y}\le(-C-1)\big[1-\frac{1}{2\pi t}\cdot\mathrm{Area}\{\boldsymbol{y}:\,|\boldsymbol{y}|<e^{C+1}\}\big]\to-C-1$$

as $t\to\infty$; the second integral does not exceed

$$\frac{1}{2\pi t}\cdot\iint\limits_{\{\boldsymbol{y}\colon|\boldsymbol{y}|<1\}}u(\boldsymbol{y})\,dy\to0\quad(t\to\infty).$$

So for t large enough the integral representing the expectation is less than $-C$: the expectation goes to $-\infty$.

31 No, it is not a martingale, otherwise we would have $E\big(u(\boldsymbol{W}_t)\big)\equiv\mathrm{const}$.

It can be proved that $u(\boldsymbol{W}_t)$ is a *supermartingale*, i. e., $-u(\boldsymbol{W}_t)$ is a submartingale.

32 Again: plain and honest differentiation.

33 Plain and honest differentiation this too; but let me show that I can differentiate. For shortness, let us introduce the notation

$$k(x^1,\,x^2,\,y)=\frac{x^2}{(x^2)^2+(x^1-y)^2};$$

we have:

$$u(x^1,\, x^2) = \pi^{-1} \int_{-\infty}^{\infty} k(x^1,\, x^2,\, y) \cdot \varphi(y)\; dy.$$

The derivative

$$\frac{\partial^j u}{(\partial x^i)^j}(x^1,\, x_2) = \pi^{-1} \int_{-\infty}^{\infty} \frac{\partial^j}{(\partial x^i)^j}\big(k(x^1,\, x^2,\, y) \cdot \varphi(y)\big)\; dy$$

(the integrals of the derivatives converge uniformly in a neighborhood of every point $(x^1,\, x^2)$ in the upper half-plane), so we have to check that

$$\frac{\partial^2 k}{(\partial x^1)^2}(x^1,\, x^2,\, y) + \frac{\partial^2 k}{(\partial x^2)^2}(x^1,\, x^2,\, y) \equiv 0. \tag{92}$$

We have:

$$\frac{\partial k}{\partial x^1}(x^1,\, x^2,\, y) = \frac{-\,2x^2(x^1 - y)}{\big((x^2)^2 + (x^1 - y)^2\big)^2},$$

$$\frac{\partial^2 k}{(\partial x^1)^2}(x^1,\, x^2,\, y) = \frac{2x^2\big(3(x^1 - y)^2 - (x^2)^2\big)}{\big((x^2)^2 + (x^1 - y)^2\big)^3},$$

$$\frac{\partial k}{\partial x^2}(x^1,\, x^2,\, y) = \frac{(x^1 - y)^2 - (x^2)^2}{\big((x^2)^2 + (x^1 - y)^2\big)^2},$$

$$\frac{\partial^2 k}{(\partial x^2)^2}(x^1,\, x^2,\, y) = \frac{2x^2\big(-\,3(x^1 - y)^2 + (x^2)^2\big)}{\big((x^2)^2 + (x^1 - y)^2\big)^3},$$

so indeed (92) is satisfied.

34 For a fixed $x \in (-\infty, \infty)$ and a positive ε, let $\delta > 0$ be such that

$$|\varphi(y) - \varphi(x)| < \varepsilon/2 \quad \text{for } |y - x| < \delta.$$

We have:

$$\begin{aligned} u(x^1,\, x^2) - \varphi(x) &= \int_{-\infty}^{\infty} \frac{\pi^{-1} \cdot x^2}{(x^2)^2 + (x^1 - y)^2} \cdot [\varphi(y) - \varphi(x)]\; dy \\ &= \int_{x-\delta}^{x+\delta} \frac{\pi^{-1} \cdot x^2}{(x^2)^2 + (x^1 - y)^2} \cdot [\varphi(y) - \varphi(x)]\; dy \\ &\quad + \int\limits_{(-\infty,\,\infty)\setminus(x-\delta,\,x+\delta)} \frac{\pi^{-1} \cdot x^2}{(x^2)^2 + (x^1 - y)^2} \cdot [\varphi(y) - \varphi(x)]\; dy \end{aligned}$$

The first integral is not greater than $\varepsilon/2$ in absolute value.

If $|x^1 - x| < \delta/2$, the second integral in the right-hand side does not exceed in absolute value

$$2 \sup_x |\varphi(x)| \cdot \int\limits_{(-\infty,\,\infty)\setminus(x-\delta,\,x+\delta)} \frac{\pi^{-1}\cdot x^2}{(x^2)^2 + (x^1 - y)^2}\, dy$$
$$= 2 \sup_x |\varphi(x)| \cdot \big[1 - \pi^{-1} \arctan\big(\frac{x + \delta - x^1}{x^2}\big) + \pi^{-1} \arctan\big(\frac{x - \delta - x^1}{x^2}\big)\big]$$
$$\le 2 \sup_x |\varphi(x)| \cdot \big[1 - \pi^{-1} \arctan\big(\frac{\delta}{2x^2}\big) + \pi^{-1} \arctan\big(\frac{-\delta}{2x^2}\big)\big] \to 0 \quad (x^2 \to 0^+).$$

So we can choose a $\delta_1 > 0$ so that the second integral is smaller than $\varepsilon/2$ for $|x^1 - x| < \delta/2$, $0 < x^2 < \delta_1$. For such x^1, x^2 the difference $u(x^1, x^2) - \varphi(x)$ is less than ε in absolute value; so the statement is proved.

35, **36**: Sine for every $c > 0$ we have $(\tau^x_{(a,\,b)})^m <$ const $\cdot\, e^{c\tau^x_{(a,\,b)}}$, and $E(e^{c\tau^x_{(a,\,b)}}) < \infty$ for $c < \pi^2/2(b-a)^2$ (see Lecture note 34), the expectations of $(\tau^x_{(a,\,b)})^m$, $m = 2, 3, 4$ (and for all other m) are finite.

37 The second moment $E\big((\tau^x_{(a,\,b)})^2\big)$ can be found as the second derivative at $t = 0$ of the moment generating function $M_{\tau^x_{(a,\,b)}}(t) = E(e^{t\tau^x_{(a,\,b)}})$, given by formulas (34.13), (34.14) for $t < \pi^2/2(b-a)^2$ (and by $M_{\tau^x_{(a,\,b)}}(t) = 1$ for $t = 0$). Better we represent the function given by formulas (34.13), (34.14) as a power series in t: for $t < 0$ the function is given by the formula

$$\frac{\sinh\big(\sqrt{-2t}\cdot(x-a)\big) + \sinh\big(\sqrt{-2t}\cdot(b-x)\big)}{\sinh\big(\sqrt{-2t}\cdot(b-a)\big)}.$$

The numerator can be represented as

$$\sqrt{-2t}\cdot(x-a) + \big(\sqrt{-2t}\cdot(x-a)\big)^3/6 + \big(\sqrt{-2t}\cdot(x-a)\big)^5/120 + \ldots$$
$$+ \sqrt{-2t}\cdot(b-x) + \big(\sqrt{-2t}\cdot(b-x)\big)^3/6 + \big(\sqrt{-2t}\cdot(b-x)\big)^5/120 + \ldots$$
$$= \sqrt{-2t}\cdot(A + Bt + Ct^2 + \ldots),$$

where

$$A = b - a, \qquad B = -\big((x-a)^3 + (b-x)^3\big)/3, \qquad C = \big((x-a)^5 + (b-x)^5\big)/30.$$

The denominator is represented as

$$\sqrt{-2t}\cdot(b-a) + \big(\sqrt{-2t}\cdot(b-a)\big)^3/6 + \big(\sqrt{-2t}\cdot(b-a)\big)^5/120 + \ldots$$
$$= \sqrt{-2t}\cdot(A + Dt + Et^2 + \ldots),$$

where A is the same, and

$$D = -(b-a)^3/3, \qquad E = (b-a)^5/30.$$

The ratio of these functions is represented by the power series

$$1 + Ft + Gt^2 + ..., \qquad F = \frac{B - D}{A}, \qquad G = \frac{C - E}{A} - \frac{D(B - D)}{A^2}.$$

It turns out that for $0 < t < \pi^2/2(b - a)^2$ the moment-generating function is represented by the same power series (of course, also for $t = 0$, where the function is equal to 1); and the second moment

$$E\big((\tau^x_{(a,\,b)})^2\big) = \frac{d^2}{dt^2} M_{\tau^x_{(a,\,b)}}(0) = 2G.$$

We can return to our formulas expressing G through A, B, C, and D, and giving expressions for these coefficients.

So indeed the second moment can be evaluated (only I did not expect the expression would be so complicated).

38 The function e^{u^2} grows at infinity faster than e^{cu}, so we have that $e^{(\tau^x_{(a,\,b)})^2} \geq \text{const} \cdot e^{c \cdot \tau^x_{(a,\,b)}}$, and $E(e^{(\tau^x_{(a,\,b)})^2}) \geq \text{const} \cdot E(e^{c \cdot \tau^x_{(a,\,b)}})$. Since for some c the expectation $E(e^{c \cdot \tau^x_{(a,\,b)}}) = \infty$ (see Lecture note 34), we get that $E(e^{(\tau^x_{(a,\,b)})^2}) = \infty$.

39 $E(\tau^{\boldsymbol{x}}_G) = m(\boldsymbol{x}) = m(x^1, x^2) = \big(R^2 - (x^1)^2 + (x^2)^2\big)/2$, $\boldsymbol{x} \in G$. It is easily checked that $m(\boldsymbol{x})$ is the solution of the boundary-value problem $\dfrac{1}{2}\,\Delta m(\boldsymbol{x}) = -1$, $\boldsymbol{x} \in G$, $m(\boldsymbol{x}) = 0$, $\boldsymbol{x} \in \partial G$.

40 $m(\boldsymbol{x}) = a^2 - (x^1)^2$.

41 Looking for $m(\boldsymbol{x})$ in the form $m(\boldsymbol{x}) = C - D \cdot (x^1)^2 - E \cdot (x^2)^2$, we get from the boundary condition that $m(\boldsymbol{x}) = C \cdot \big(1 - A \cdot (x^1)^2 - B \cdot (x^2)^2\big)$; now we differentiate: $\dfrac{1}{2}\,\Delta m(\boldsymbol{x}) = -C \cdot (A + B)$, which should be equal to -1. So $C = 1/(A + B)$, $E(\tau^{\boldsymbol{x}}_G) = \dfrac{1 - A \cdot (x^1)^2 - B \cdot (x^2)^2}{A + B}$.

42 The operator $\mathfrak{L}$ is given by

$$\mathfrak{L}f(x) = \begin{cases} \frac{1}{2}\,f''(x) + f'(x), & f'(x) \geq 1, \\ \frac{1}{2}\,f''(x) + 1, & -1 \leq f'(x) \leq 1, \\ \frac{1}{2}\,f''(x) - f'(x), & f'(x) \leq -1. \end{cases} \tag{93}$$

Of course the solution of the boundary-value problem $\mathfrak{L}v(x) = 0$, $x \in (-c, c)$, $v(-c) = v(c) = 0$ should be symmetric with respect to the point $x = 0$. Since $\frac{1}{2}\,v(x) \leq -1 < 0$, the first derivative $v'(x)$ is decreasing in the interval, and $v(x)$ should satisfy an equation with the operator given by the first line in (93) in an interval from $-c$ to some $-d$, another one with the operator as in the second line in an interval from $-d$ to d near the middle of the interval $(-c, c)$, and with the operator in the third line near its right end. The simplest one is the one in the middle:

$$\tfrac{1}{2} v''(x) + 1 = 0,$$

its symmetric solution being

$$v(x) = C - x^2, \qquad C = \text{const.} \tag{94}$$

This function satisfies the inequalities $-1 \le v'(x) \le 1$ in the interval $[-1/2, 1/2]$, so we see that the small interval $[-d, d]$ should be $[-1/2, 1/2]$.

This is clearly impossible if $c < 1/2$; but for these c, and for $c = 1/2$, the solution is given by one formula (94), with C determined from the boundary conditions: $C = c^2$,

$$v(x) = c^2 - x^2, \qquad -c \le x \le c.$$

For $c > 1/2$ we have to solve the equation

$$\tfrac{1}{2} v''(x) + v'(x) = 0$$

in the interval from $-c$ to $-1/2$, with boundary conditions

$$v(-c) = 0, \qquad v'\big((-1/2)^-\big) = 1.$$

It is easily checked that

$$v(x) = (e^{2c-1} - e^{-2x-1})/2, \qquad -c \le x \le -1/2.$$

From the condition $v\big((-1/2)^-\big) = v\big((-1/2)^+\big)$ we find the constant C in (94): $C = \frac{1}{2}(e^{2c-1} - 1) + 1/4$; and we can write the solution in the whole interval $[-c, c]$, including its part from $1/2$ to c:

$$v(x) = \begin{cases} \frac{1}{2}(e^{2c-1} - e^{-2x-1}), & -c \le x \le -1/2, \\ \frac{1}{2}(e^{2c-1} - 1) + 1/4 - x^2, & -1/2 \le x \le 1/2, \\ \frac{1}{2}(e^{2c-1} - e^{2x-1}), & 1/2 \le x \le c. \end{cases}$$

The optimal strategy is, for the process $X_t^{x,u}$ being at a point $x \in (-c, c)$, to use the controlling parameter $u = \hat{u}(x) \in [-1, 1]$ at which the maximum

$$\max_{u \in [-1, 1]} [\tfrac{1}{2} v''(x) + u \cdot v'(x) + 1 - |u|]$$

is reached. So for $c \le 1$ the optimal control is

$$\hat{u}(x) \equiv 0$$

since $|v(x)| < 1$ (draw the graph of $\frac{1}{2} v''(x) + u \cdot v'(x) + 1 - |u|$ as a function of u). For $c > 1/2$ we have:

$$\hat{u}(x) = \begin{cases} 1, & x < -1/2, \\ 0, & -1/2 < x < 1/2, \\ -1, & x > 1/2. \end{cases}$$

As for the control that we should use *at* the points $x = -1/2,\ 1/2$, the maximum is reached in a whole interval $u \in [0, 1]$, correspondingly, $u \in [-1, 0]$; but what control we are using at these two points is not important, because the time spent by our process at these points is almost surely 0.