I.

Очевидно, что для любого конкретного числа вероятность встретить именно его равна 0, так как отрезок, состоящий из этого числа, имеет длину 0. Можно нестрого заметить, что вероятность того, что произвольное число из массива окажется на каком-то отрезке, пропорциональна длине этого отрезка, так как, например, если разбить отрезок (0, 1) на две половины, вероятность получить число на каждой из половин будет одинакова и равна половине вероятности получить число на (0, 1) и равна 0.5 (при этом вероятность получить число ровно 0.5, являющееся серединой отрезка (0, 1), равна 0). Каждую из половин можно разбивать дальше и получить разбиение массива на очень маленькие отрезки, при этом вероятность получить число на произвольном отрезке будет пропорциональна количеству этих маленьких отрезков в нём и практически пропорциональна длине.

Посмотрим на крайние значения приведённых функций A-E. Так как не говорится, что массив маленький, а функции непрерывны на (0, 1), наверно, все функции практически достигнут и своих крайних значений на отрезке (0, 1). Пусть операция "^" означает возведение в степень. Также оценим медианы. Для медианы массива значений каждой функции будет выполняться, что при равновероятном выборе одного значения этой функций из массива её значений вероятность получить значение меньше медианы будет пропорциональна длине отрезка, равной 0.5, как и вероятность получить значение больше.

A) -x^2+x+0.5 = -(x^2-x-0.5) = -((x-0.5)^2 - 0.25 - 0.5) = -((x-0.5)^2 - 0.75) = 0.75 - (x-0.5)^2.

Значение зависит только от |x-0.5|. |x-0.5| достигает значения от 0.5 (для x ~=(примерно равно) 0 или 1) до 0 (для x ~= 0.5). Тогда значение функции f(x) будет от 0.75 до 0.75-0.5^2=0.75-0.25=0.5.

B) f(x) = 1-x. Для x от 0 до 1 значения f(x) будут от 1 до 0, причём распределение очевидно будет равномерным, а его медиана останется на месте и будет равна 0.5. (Так как вероятность получить f(x) <= 0.5 зависит от длины отрезка (0, 0.5), а это 0.5, а вероятность получить f(x) >= 0.5 тоже зависит от длины отрезка, равной 0.5 -- длины отрезка (0.5, 1).)

C) Функция 2^(x-1) монотонна. Её крайние значения для x на (0,1) - это 2^(0-1) = 1/2 = 0.5 и 2^(1-1) = 1.

D) Функция x^2 монотонна на неотрицательных числах. Её крайними значениями будут 0^2=0 и 1^2=1. Но при этом для всех чисел от 0 до 1, кроме 0 и 1, x^2 < x, так как 0 < x < 1 (домножим на x, получим 0 < x^2 < x). Тогда медиана будет меньше, чем медиана равномерного распределения на (0, 1), то есть меньше 0.5.

E) Функция x/2 + 0.5 монотонна. Её минимум для x на (0, 1) равен 0/2+0.5 = 0.5, максимум -- 1/2 + 0.5 = 1. При этом медиана массива её значений равна 0.75, так как вероятность получить значение f(x), лежащее на (0.5, 0.75), пропорциональна длине отрезка (0, 0.5), равной 0.5, и вероятность получить значение на (0.75, 1), пропорциональна длине отрезка (0.5, 1), равной 0.5.

Получаем приблизительные отрезки значений: функции A - от 0.5 до 0.75, функции B - от 0 до 1, причём медиана 0.5, функции C - от 0.5 до 1, функции D - от 0 до 1, причём медиана меньше 0.5, функции E - от 0.5 до 1, причём медиана 0.75.

Найдём график распределения значений функции A.

Определим, какая медиана должна быть. Для равномерного распределения x на (0,1), (x-0.5) равномерно распределено на (-0.5, 0.5).

Тогда |x-0.5| равномерно распределено на (0, 0.5) (так как |x-0.5| равномерно распределено на (0, 0.5) и для x на (-0.5, 0), и для x на (0, 0.5), а значит и для объединения (-0.5, 0) и (0, 0.5) (так как они пересекаются всего по одному числу 0)).

Найдём медиану распределения. Так как при возрастании |x-0.5| (x-0.5)^2=|x-0.5|^2 возрастает, медиана будет равна значению (x-0.5)^2 при |x-0.5|, являющемся медианой множества своих возможных значений. Для равномерного распределения это середина, т.е. получаем медиану (x-0.5)^2 при |x-0.5|=(0+0.5)/2=0.25, тогда медиана равна (0.25-0.5)^2 =

$0.25^2 = 1/16$. Тогда медиана значений $-(x-0.5)^2$ для x из массива значений x будет (примерно) равна минус медиане $(x-0.5)^2$ и равна -1/16, а медиана $0.75-(x-0.5)^2$ -- 3/4 - 1/16 = 12/16-1/16=11/16 = 0.6875.
Таким образом, медиана окажется больше 0.6. Тогда среди ящиков 1 и 4, каждый из которых имеет минимум и максимум примерно 0.5 и 0.75, для функции A подходит ящик 4, так как у второго ящика медиана меньше 0.6.

Для функции B по значениям миинмума, максимума и медианы подходит только ящик 5, так как он показывает минимум примерно 0, максимум примерно 1, медиану примерно 0.5, а похожий на него ящик 6 показывает другую медиану.

Перед тем, как определить ящик для функции C, определим ящик для функции E. Для функции E подходит только ящик 3, так как он показывает минимум примерно 0.5, максимум примерно 1, и медиану примерно 0.75. Похожий на него ящик 2 показывает другую медиану.

Теперь можно определить ящик для функции C. Среди ящиков с минимумом 0.5 и максимумом 1 остаётся только ящик 2. Значит, ящик 2 показывает распределение значений функции C. Заметим, что в условии сказано, что ящики построены на полученных массивах, поэтому можно пользоваться "методом исключения".

Для функции D по значениям минимума и максимума подходят ящики 5 и 6. Но мы определили, что ящик 5 показывает значения функции B, тогда значения функции D показывает ящик 6, тем более, что медиана, которую показывает ящик, действительно меньше 0.5, как и должно быть для функции D.

Тогда получаем такое соответствие: A-4, B-5, E-3, C-2, D-6.

II.
На приведённом графике, показывающем распределение зарплат, можно увидеть, что по результатам опроса:
Есть относительно резкий пик распределения зарплат в районе 16000-26000 рублей с максимумом количества для зарплаты 20000 рублей, а также большое число людей, зарплата которых от 35000 до 65000 рублей, причём среди этой группы наиболее популярная зарплата примерно 48000 и чем дальше зарплата от популярной, тем меньше число людей, получающих такую зарплату. Для зарплат вне двух указанных групп число людей, их получающих, в среднем меньше. Среди людей, получающих зарплату меньше 33000 рублей, опрос прошла только половина. Можно считать, что зарплаты не прошедших опрос так же распределены, как и меньшие 33000 зарплаты людей, прошедших опрос. Таким образом, на самом деле пик распределения ещё примерно в 2 раза выше и все зарплаты в районе 16000-26000 рублей получают примерно в 2 раза больше людей, а людей с зарплатой вне двух популярных отрезков (16000-26000 и 35000-65000) всё равно мало. Это отвечает на вопрос "Опиши, как истинное распределение зарплат будет отличаться от представленного выше и почему."
Средняя зарплата в таблице завышена, потому что мы, используя результаты опроса, не учли многих людей с зарплатой ниже посчитанной нами средней, но учли почти всех людей с зарплатой выше посчитанной нами средней: мы учли мало людей с зарплатой 16000-26000, но почти всех с зарплатой 35000-65000. Даже если мы не учли каких-то людей с очень большой зарплатой, мы оцениваем зарплаты настолько большого числа людей, что это не очень сильно повлияет на среднее.
Медианная зарплата, записанная в таблице, находится в области зарплат, которые получают мало людей. Медианная зарплата в таблице завышена, потому что мы, используя результаты опроса, не учли многих людей с зарплатой ниже посчитанной нами медианной, но учли

почти всех людей с зарплатой выше посчитанной нами медианной: мы учли мало людей с зарплатой 16000-26000, но почти всех с зарплатой 35000-65000. Медиана менее чувствительна к выбросам, поэтому то, что у не учтённых людей может быть очень большая зарплата, не сильно влияет на смещение медианы.
75-ый перцентиль уменьшится, так как практически все не учтённые зарплаты ниже нынешнего (записанного в таблице)75-го перцентиля. Тогда после их учёта значение, являющееся 75-ым перцентилем, станет большим перцентилем; следовательно, настоящий 75ый перцентиль станет ниже.
25-ый перцентиль уменьшится: если учесть зарплаты сотрудников, не учтённые по результатам опроса, то примерно половина не учтённых зарплат окажется меньше нынешнего 25-го перцентиля, и примерно половина - больше, так как распределение зарплат в районе 16000-26000 примерно симметрично относительно 20000 ~= 19979 -- 25-ый перцентиль. Тогда нынешнее число-25ый перцентиль станет большим перцентилем (ближе к 50му перцентилю); следовательно, настоящий 25-ый перцентиль уменьшится.
Максимальная зарплата может быть верной или заниженной. Мы видим, что наибольшие зарплаты по результатам опроса получает очень мало человек: если ещё большие зарплаты тоже получают мало человек, они все могли не проголосовать, и тогда максимальная зарплата занижена. Даже если проголосовали почти все сотрудники, неизвестно, проголосовали ли почти все с зарплатой 90 тысяч и больше, так как таких людей, судя по всему, очень мало, и даже если бы они не проголосовали, это не сделало бы утверждение руководителя HR-департамента неверным. Наверно, максимальная зарплата немного занижена.