Tutorial Worksheet 2 - Reviews of Probability Distributions (with Solutions)

Objective Questions

1. Which of the following mentioned standard Probability density functions is applicable to discrete Random Variables?

   (a) Gaussian Distribution (b) Poisson Distribution (c) Gamma Distribution (d) Exponential Distribution

   Solution: Poisson Distribution

2. If the values taken by a random variable are negative, the negative values will have

   (a) Positive probability (b) Negative probability (c) May have negative or positive probabilities (d) Insufficient data

   Solution: Positive probability

3. The expected value of a random variable is its

   (a) Mean (b) Standard Deviation (c) Mean Deviation (d) Variance

   Solution: Mean

4. In a Binomial Distribution, if 'n' is the number of trials and 'p' is the probability of success, then the mean and variance is given by

   (a) $np, np(1-p)$ (b) $np(1-p), np$ (c) n, np^2 (d) p, np

   Solution: $np, np(1-p)$

5. For larger values of 'n', Binomial Distribution

   (a) loses its discreteness (b) tends to Poisson Distribution (c) stays as it is (d) gives oscillatory values

   Solution: tends to Poisson Distribution

6. The recurrence relation between $P(x)$ and $P(x+1)$ in a Poisson distribution with parameter m is given by

   (a) $P(x+1) - mP(x) = 0$ (b) $mP(x+1) - P(x) = 0$ (c) $(x+1)\,\mathrm{P}(x+1) - m\,\mathrm{P}(x) = 0$ (d) $(x+1)P(x) - xP(x+1) = 0$

   Solution: $(x+1)\,\mathrm{P}(x+1) - m\,\mathrm{P}(x) = 0$

7. Which of the following statement sounds reasonable?

   (a) If the sample size increases sampling distribution approaches normal distribution.
   (b) If the sample size decreases then the sample distribution approaches normal distribution.
   (c) If the sample size increases then the sampling distribution approaches an exponential distribution.
   (d) If the sample size decreases then the sampling distribution approaches an exponential distribution.

   Solution: If the sample size increases sampling distribution approaches normal distribution.

8. If μ and σ denote the mean and standard deviation of a population, then the standard normal distribution is better described as (select the correct option from the list of options given below):

   (a) $f(x;\mu,\sigma) = \begin{cases} \frac{1}{\mu-\sigma}, & \mu \le x \le \sigma \\ 0, & \text{Otherwise} \end{cases}$ (b) $f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$

   (c) $f(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty$ (d) $f(x;\mu,\sigma) = \begin{cases} \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2}, & x \ge 0 x < 0 \\ 0, & \text{otherwise} \end{cases}$

Solution: $f(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty$

9. If $f(x)$ is a probability density function of any continuous random variable, then which of the follow statement(s) is (are) correct?

   (a) $0 \leq f(x) \leq 1$ (b) $P(a \leq X \leq b) = \int_b^a f(x)dx < 1$ (c) $y = \int_{-\alpha}^{\alpha} xf(x)dx$ there exist $y \in R$ (d) $z = \int_{-\alpha}^{\alpha} (x-\mu)^2 f(x)dx$ there exist $\mu \in R$ and $z \in R$

   Solution: $0 \leq f(x) \leq 1$ and $P(a \leq X \leq b) = \int_b^a f(x)dx < 1$

10. In the following Table, Column A lists some sampling distributions, whereas Column B lists the name of sampling distributions. All symbols bear their usual meanings.

Column A		Column B	
(A)	$\frac{X-\mu}{\sigma}$	(W)	Normal distibution
(B)	$\frac{X-\mu}{s/\sqrt{n}}$	(X)	Chi-squared distribution
(C)	$\frac{S_1/d_1}{S_2/d_2}$	(Y)	t-distribution
(D)	$\sum_{i=1}^{k} Z_i^2$	(Z)	F distribution

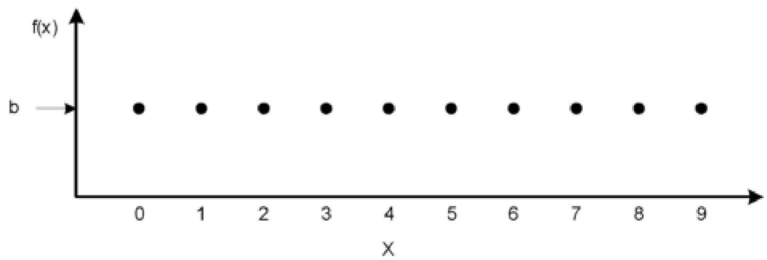
Some matchings from Column A and Column B are given below. Select the correct matching?

(a) (A)-(Z), (B)-(X), (C)-(Y), (D)-(W) (b) (A)-(X), (B)-(Z), (C)-(Y), (D)-(W) (c) (A)-(Y), (B)-(W), (C)-(X), (D)-(Z) (d) (A)-(W), (B)-(Y), (C)-(Z), (D)-(X)

Solution: (A)-(W), (B)-(Y), (C)-(Z), (D)-(X)

Subjective Questions

Problem 1. A bitcoin if you toss it gives any value in the range $[0, \ldots, 9]$ both inclusive. Assume that the random variable X represents the toss value of the bitcoin which has a discrete uniform distribution and is shown in the following figure.



(a) What is $P(X = x_i) = f(x_i)$ for any value of $x_i \in [0 \ldots 9]$?

(b) What is the value of b (as marked in the figure)?

(c) Calculate the mean μ of this distribution.

(d) Calculate the variance σ^2 of this distribution.

Solution:

(a) $P(X = x_i) = f(x_i) = \frac{1}{10} = 0.1$

(b) $b = f(x_i) = \frac{1}{10} = 0.1$

(c) $\mu = \sum_1^{10} x_i \cdot f(x_i) = \frac{b+a}{2} = \frac{9+0}{2} = 4.5$

(d) $\sigma^2 = \sum_1^{10} (x_i - \mu)^2 \cdot f(x_i) = \frac{(b-a+1)^2-1}{12} = \frac{10^2-1}{12} = 8.25$

Problem 2. A quiz test for a course Data Analytics was conducted for a total score of 100 where 600 students took the test. From the result of the test it was found that mean score $\mu = 90$ and standard deviation $\sigma = 20$. Students are randomly distributed among six sections and each section includes 100 students. In one of the section of 100 students, the mean score is found as 86.

(a) What is the standard error rate?

(b) If you select any section at random, what is the probability of getting a mean score is 86 or lower?

Solution:

(a) As per the Central Limit Theorem, the standard error is $\varepsilon = \frac{\sigma}{\sqrt{n}} = \frac{20}{10} = 2.0$.

(b) The sample distribution statistics can be obtained with the z-distribution. For the sample, The probability of getting 86 or lower is $P(Z < -2.0)$. From the standard normal distribution table it is found that $P(Z < -2.0) = 0.0228$.

Problem 3. On average, 3 traffic accidents per month occur at a certain intersection. What is the probability that in any given month at this intersection

(a) exactly 5 accidents will occur?

(b) fewer than 3 accidents will occur?

(c) at least 2 accidents will occur?

Solution:

(a) $P(X=5) = \frac{e^{-3} \cdot 3^5}{5!} = 0.1008$.

(b) $P(X<3) = P(X \leq 2) = e^{-3} \cdot \sum_{x=0}^{2} \frac{3^x}{x!} = 0.4232$.

(c) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - e^{-3} \cdot \sum_{x=0}^{1} \frac{3^x}{x!} = 0.8009$.

Problem 4. A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates that the defective rate of the device is 3%.

(a) The inspector randomly picks 20 items from a shipment. What is the probability that there will be at least one defective item among these 20?

(b) Suppose that the retailer receives 10 shipments in a month and the inspector randomly tests 20 devices per shipment. What is the probability that there will be exactly 3 shipments each containing at least one defective device among the 20 that are selected and tested from the shipment?

Solution:

(a) Denote by X the number of defective devices among the 20. Then X follows a $b(x; 20, 0.03)$ distribution. Hence, $P(X \geq 1) = 1 - P(X=0) = 1 - b(0; 20, 0.03) = 1 - (0.03)^0(1-0.03)^{20-0} = 0.4562$.

(b) In this case, each shipment can either contain at least one defective item or not. Hence, testing of each shipment can be viewed as a Bernoulli trial with $p = 0.4562$ from part (a). Assuming independence from shipment to shipment and denoting by Y the number of shipments containing at least one defective item, Y follows another binomial distribution $b(y; 10, 0.4562)$. Therefore,

$$P(Y=3) = \left(\begin{array}{c} 10 \\ 3 \end{array} \right) 0.4562^3(1-0.4562)^7 = 0.1602$$

Problem 5. A soft-drink machine is regulated so that it discharges an average of 200 milliliters per cup. If the amount of drink is normally distributed with a standard deviation equal to 15 milliliters,

(a) What fraction of the cups will contain more than 224 milliliters?

(b) What is the probability that a cup contains between 191 and 209 milliliters?

(c) How many cups will probably overflow if 230- milliliter cups are used for the next 1000 drinks?

Solution:

(a) $z = \frac{224-200}{15} = 1.6$ Fraction of the cups containing more than 224 millimeters is $P(Z > 1.6) = 0.0548$

(b) $z_1 = \frac{191-200}{15} = -0.6, z_2 = \frac{209-200}{15} = 0.6;$

$$P(191 < X < 209) = P(-0.6 < Z < 0.6) = 0.7257 - 0.2743 = 0.4514$$

(c) $z = \frac{230-200}{15} = 2.0; P(X > 230) = P(Z > 2.0) = 0.0228$. Therefore, $(1000)(0.0228) = 22.8$ or approximately 23 cups will overflow.

Problem 6. The Los Angeles Times (December 13, 1992) reported that what airline passengers like to do most on long flights is rest or sleep; in a survey of 3697 passengers, almost 80% did so. Suppose that for a particular route the actual percentage is exactly 80%, and consider randomly selecting six passengers.

(a) Calculate $p(4)$, and interpret this probability.

(b) Calculate $p(6)$, the probability that all six selected passengers rested or slept.

(c) Determine $P(X \geq 4)$.

Solution: Suppose that the actual percentage of passengers who rested or slept for a particular route is exactly 80%. Let x denote the number of six randomly selected passengers who rested or slept, so x is a binomial random variable with $n = 6$ and $p = 0.8$. Then the probability mass function of x is given by:

$$\begin{aligned} p(x) = P(x \text{ successes among } n \text{ trials}) &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n \\ &= \frac{6!}{x!(6-x)!} (0.8)^x (1-0.8)^{6-x}, \quad x = 0, 1, 2, \ldots, 6 \end{aligned}$$

(a) Now find $p(4)$ using the expression of probability mass function given above.

$$\begin{aligned} p(4) &= \frac{6!}{4!(6-4)!} \times (0.8)^4 \times (1-0.8)^{6-4} = \left(\frac{720}{24 \times 2}\right) \times (0.8)^4 \times (0.2)^2 \\ &= (15) \times (0.4096) \times (0.04) = 0.2458 \end{aligned}$$

If a group of 6 passengers is examined, the long-run percentage of exactly four of those passengers rested or slept will be 24.58%.

(b) Now find the probability that all six selected passengers rested or slept:

$$p(6) = \frac{6!}{6!(6-6)!} \times (0.8)^6 \times (1-0.8)^{6-6} = \left(\frac{720}{720 \times 1}\right) \cdot (0.8)^6 \cdot (0.2)^0 = 0.2621$$

If group after group of 6 passengers is examined, the long-run percentage of all six of those passengers rested or slept will be 26.21%.

(c) Now find $P(x \geq 4)$, the probability that at least 4 of six selected passengers rested or slept:

$$\begin{aligned} P(x \geq 4) &= p(4) + p(5) + p(6) \\ &= \frac{6!}{4!2!} \times (0.8)^4 \times (1-0.8)^2 + \frac{6!}{5!1!} \times (0.8)^5 \times (1-0.8) + \frac{6!}{6!0!} \times (0.8)^6 \times (1-0.8)^0 \\ &= 0.2458 + 0.3932 + 0.2621 = 0.9011 \end{aligned}$$

If group after group of 6 passengers is examined, the long-run percentage of at least four of those passengers rested or slept will be 90.11%.

Problem 7. Suppose that the number of telephone calls coming into a telephone exchange between 10 A.M and 11 A.M. say, X_1 is a random variable with Poisson distribution with parameter 2. Similarly the number of calls arriving between 11A.M. to 12 noon, say, X_2 has a Poisson distribution with parameter 6. If X_1 and X_2 are independent, what is the probability that more than 5 calls come in between 10 A.M. and 12 noon?

Solution:

We are given, $X_1 = P(2)$ and $X_2 = P(6)$. Let, $X = X_1 + X_2$. By the additive property of Poisson distribution, X is also a Poisson variable with parameter $2 + 6 = 8$. Hence, the probability of x calls in between 10 A.M. to 12 noon is given by, $P(X = x) = \frac{e^{-8}8^x}{x!}; x = 0, 1, 2, \cdots$.

Probability that more than 5 calls come in between 10 A.M. to 12 noon is given by:

$$P(X > x) = 1 - P(X \leq 5) = 1 - \sum_{x=0}^{5} \frac{e^{-8}8^x}{x!} = 0.80876$$

Problem 8. Suppose a train arrives at a subway station regularly every 20 min. If a passenger arrives at the station without knowing the timetable, then find the probability that the man will have to wait at least 10 min ? What is the average waiting time?

Solution:

Let, X be the random variable denoting the waiting time of the passenger. Therefore, $X \sim$ Uniform $(0, 20)$

The probability density function of X is given by:

$$f(x) = \begin{cases} \frac{1}{20}, & 0 < x < 20 \\ 0, & x > 20 \end{cases}$$

The probability that the man will have to wait at least 10 min

$$\begin{aligned} P(X \geq 10) &= 1 - P(X < 10) \\ &= 1 - \int_0^{10} \frac{1}{20} dx \\ &= 1 - \frac{10-0}{20} = 0.5 \end{aligned}$$

The average waiting time is given by: $E(X) = \frac{0+20}{2} = 10$

Problem 9. There are 600 data science students in the undergraduate classes of a university, and the probability for any student to need a copy of a particular book from the university library on any day is 0.05. How many copies of the book should be kept in the university library so that the probability may be greater than 0.90 that none of the students needing a copy from the library has to come back disappointed?

Solution:

We are given: $n = 600; p = 0.05 \Rightarrow \mu = n \times p = 600 \times 0.05 = 30$ and $\sigma^2 = n \times p \times (1-p) = 600 \times 0.05 \times (1-0.05) = 28.5$, i.e., $\sigma = \sqrt{28.5} = 5.34$.
We want x such that, $P(X < x) > 0.90 \Rightarrow P\left(\frac{X-30}{5.34} < \frac{x-30}{5.34}\right) > 0.90 \Rightarrow P\left(Z < \frac{x-30}{5.34}\right) > 0.90 \Rightarrow \frac{x-30}{5.34} = 1.28$ (Using Standard Normal Table). $\Rightarrow x \simeq 37$
Hence, the university should keep at least 37 copies of the book.

Problem 10. Emissions of nitrogen oxides, which are major constituents of smog, can be modeled using a normal distribution. Let X denote the amount of this pollutant emitted by a randomly selected vehicle (in parts per billion). The distribution of X can be described by a normal distribution with mean of 1.6 and standard deviation of 0.4. Suppose that the EPA wants to offer some sort of incentive to get the worst polluters off the road. What emission levels constitute the worst 10% of the vehicles?

Solution:

Given that, mean $\mu = 1.6$; standard deviation $\sigma = 0.4$, using Standard Normal Table we have, $P(Z > z) = 0.10 \Rightarrow 1 - P(Z < z) = 0.10 \Rightarrow P(Z < z) = 0.90 \Rightarrow z = 1.282$.
Using z-score formula we have, $x = z \times \sigma + \mu \Rightarrow x = 1.282 \times 0.4 + 1.6 \Rightarrow x = 2.1128$. The worst 10% of vehicles are those with emission levels greater than 2.1128 parts per billion.

Problem 11. A symmetric die is thrown 600 times. Find the lower bound for the probability of getting 80 to 120 sixes.

Solution:

Let S be the total number of successes. Then,

$$E(S) = n \times p = 600 \times \frac{1}{6} = 100 \text{ and } V(S) = n \times p \times q = 600 \times \frac{1}{6} \times \frac{5}{6} = \frac{500}{6}$$

Using Chebychev's inequality, we get

$$P[|S - E(S)| < k\sigma] \geq 1 - \frac{1}{k^2} \Rightarrow \mathrm{P}\{|\mathrm{S} - 100| < \mathrm{k}\sqrt{500/6}\} \geq 1 - \frac{1}{\mathrm{k}^2}$$

Taking $\mathrm{k} = -\frac{20}{\sqrt{500/6}}$, we get

$$P(80 \leq S \leq 120) \geq 1 - \frac{1}{400 \times (6/500)} = \frac{19}{24}$$

Problem 12. Let $f(x) = \begin{cases} \frac{2}{3}x, & 1 \leq x \leq 2 \\ 0, & \text{otherwise.} \end{cases}$ Give a bound using Chebyshev's inequality for $P\left(\frac{10}{9} \leq X \leq 2\right)$. Calculate the actual probability. How do they compare?

Solution:

The mean and the variance are

$$E(X) = \int_1^2 x\frac{2}{3}xdx = \frac{14}{9}, \quad \text{Var}(X) = \int_1^2 x^2\frac{2}{3}xdx - E(X)^2 = 0.080247.$$

Thus, $P\left(\frac{10}{9} \leq X \leq 2\right) = P\left(\frac{14}{9} - \frac{4}{9} \leq X \leq \frac{14}{9} + \frac{4}{9}\right) = P\left(-\frac{4}{9} \leq X - \frac{14}{9} \leq +\frac{4}{9}\right) = P\left(\left|X - \frac{14}{9}\right| \leq \frac{4}{9}\right)$.
By Chebyshev's inequality we know $P(|X-\mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}$. Hence, here we have $k\sigma = \frac{4}{9} \Rightarrow \text{k} = \frac{4}{9\sqrt{0.080247}}$. Thus the required bound is 0.59375
Actual probability is $P\left(\frac{10}{9} \leq X \leq 2\right) = \int_{\frac{10}{9}}^{2} \frac{2}{3}xdx = 0.92181$, which is much more than the lower bound.

Problem 13. A student took two national aptitude tests. The national average and standard deviation were 475 and 100, respectively, for the first test and 30 and 8, respectively, for the second test. The student scored 625 on the first test and 45 on the second test. Use **z** scores to determine on which exam the student performed better relative to the other test takers.

Solution:

The formula for z score is $z = \frac{x-\text{mean}}{\text{standard deviation}}$.
For the first exam $z = \frac{625-475}{100} = 1.5$ and for the second exam $z = \frac{45-30}{8} = 1.875$
By looking at the probabilities in a z score table we have $\text{P}(x < z) = \begin{cases} 0.93319, & z = 1.5 \\ 0.9696, & z = 1.875. \end{cases}$
From the above probabilities we can say that the student performed better than 93.32% of others on the first test and better than 96.96% of others on the second.
Hence, the student performed better relative to others on the second exam.

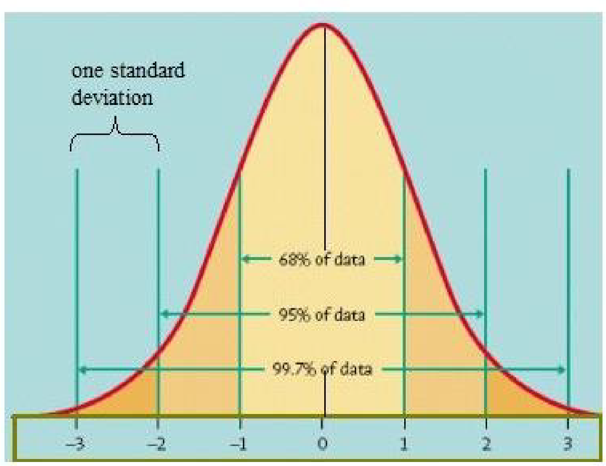
Problem 14. A sample of concrete specimens of a certain type is selected, and the compressive strength of each specimen is determined. The mean and standard deviation are calculated as $\bar{x} = 3000$ and $s = 500$, and the sample histogram is found to be well approximated by a normal curve.

(a) Approximately what percentage of the sample observations are between 2500 and 3500?

(b) Approximately what percentage of sample observations are outside the interval from 2000 to 4000 ?

(c) What can be said about the approximate percentage of observations between 2000 and 2500 ?

(d) Why would you not use Chebyshev's Rule to answer the questions posed in Parts (a)-(c)?



Solution:

(a) By the empirical rule, we know that 68% lies within 1 standard deviation from the mean and thus between 2500 and 3500.

(b) By the empirical rule, we know that 95% lies within 2 standard deviations from the mean and thus between 2000 and 4000, thus 5% lies outside the given interval.

(c) Since the normal distribution is symmetric, we know that

$$\frac{95\% - 68\%}{2} = 13.5\%$$

has to lie between 2000 and 2500.

(d) You should not use Chebyshev's Rule, because the empirical rule gives more precise estimations and we can use the empirical rule if the distribution is approximately normal.

Problem 15. An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

Solution:

Given that, $\mu = 800, \sigma = 40, n = 16$,

$$\Rightarrow P(x < 775) = P\left(Z < \frac{775 - 800}{\frac{40}{\sqrt{16}}}\right)$$
$$\Rightarrow P(Z < -2.5) = 0.5 - P(0 < Z < 2.5) = 0.5 - 0.4938 = 0.0062.$$

Problem 16. Two independent experiments are run in which two different types of paints are compared eighteen specimens are painted using type a and the drying time in houses is recorded for each the same is done with type b. The population standard deviations are both known to be 1.0, assuming that the mean drying time is equal for the two types of paints, find $P(x_a - x_b) > 1.0)$, where x_a and x_b are average drying times for sample size $n_a = n_b = 18$

Solution:

From the sampling distribution of $x_a - x_b$, we know that the distribution is approximately normal with mean

$$\mu_{x_a - x_b} = \mu_a - \mu_b = 0$$

and variance

$$\sigma^2_{x_a - x_b} = \frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b} = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}.$$

Corresponding to the value $x_a - x_b = 1.0$, we have

$$z = \frac{1 - (\mu_a - \mu_b)}{\sqrt{1/9}} = \frac{1 - 0}{\sqrt{1/9}} = 3.0.$$

So

$$P(x_a - x_b > 1.0) = P(z > 3.0) = 1 - P(z < 3.0) = 1 - 0.998650 = 0.001350.$$

Problem 17. A manufacturer of car batteries guarantees that the batteries will last, on average, 3 years with a standard deviation of 1 year. If five of these batteries have lifetimes of $1.9, 2.4, 3.0, 3.5$, and 4.2 years, should the manufacturer still be convinced that the batteries have a standard deviation of 1 year? Assume that the battery lifetime follows a normal distribution.

Solution:

$$s^2 = \frac{n\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2}{n(n-1)} = \frac{5 * 48.26 - 15^2}{5 * 4} = 0.815$$
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{4 * 0.815}{1} = 3.26$$

Since $n = 5, \chi^2$ has $\nu = n - 1 = 4$ degrees of freedom. From χ^2 Table with $\nu = 4$, wee see that $\chi^2_{0.025} = 11.143$ and $\chi^2_{0.975} = 0.484$.

Since 95% of the values with 4 degrees of freedom fall between 0.484 and 11.143, the computed value with $\sigma^2 = 1$ is reasonable (since $\chi^2 = 3.26$ falls within this range). Therefore the manufacturer has no reason to suspect that the standard deviation is other than 1 year.