Sums and Differences of Random Variables

39.3

Introduction

In some situations, it is possible to easily describe a problem in terms of sums and differences of random variables. Consider a typical situation in which shafts are fitted to cylindrical sleeves. One random variable is used to describe the variability of the diameter of the shaft, and one is used to describe the variability of the sleeves. Clearly, we need to know how the total variability involved affects the fitting of shafts and sleeves. In this Section, we will confine ourselves to cases where the random variables are normally distributed and independent.

Prerequisites

Before starting this Section you should ...

- be familiar with the results and concepts met in the study of probability
- be familiar with the normal distribution

Learning Outcomes

On completion you should be able to ...

- describe a variety of problems in terms of sums and differences of normal random variables
- solve problems described in terms of sums and differences of normal random variables

1. Sums and differences of random variables

In some situations, we can specify a problem in terms of sums and differences of random variables. Here we confine ourselves to cases where the random variables are normally distributed. Typical situations may be understood by considering the following problems.

Problem 1

In a certain mass-produced assembly, a 3 cm shaft must slide into a cylindrical sleeve. Shafts are manufactured whose diameter S follows a normal distribution $S \sim N(3, 0.004^2)$ and cylindrical sleeves are manufactured whose internal diameter C follows a normal distribution $C \sim N(3.010, 0.003^2)$. Assembly is performed by selecting a shaft and a cylindrical sleeve at random. In what proportion of cases will it be impossible to fit the selected shaft and cylindrical sleeve together?

Discussion

Clearly, the shaft and cylindrical sleeve will fit together only if the diameter of the shaft is smaller than the internal diameter of the cylindrical sleeve. We need the difference of the two random variables C and S to be greater than zero. We can take the difference $C - S$ and find its distribution. Once we do this we can then ask the question "What is the probability that the inside diameter of the cylindrical sleeve is greater than the outside diameter of the shaft, i.e. what is P$(C - S > 0)$?" Essentially we are trying to ensure that the internal diameter of the cylindrical sleeve is larger than the external diameter of the shaft.

Problem 2

A manufacturer produces boxes of woodscrews containing a variety of sizes for a local DIY store. The weight W (in kilograms) of boxes of woodscrews manufactured is a normal random variable following the distribution $W \sim N(1.01, 0.004)$. Note that 0.004 is the *variance*. Find the probability that a customer who selects two boxes of screws at random finds that their combined weight is greater than 2.03 kilograms.

Discussion

In this problem we are looking at the effects of adding two random variables together. Since all boxes are assumed to have weights W which follow the distribution $W \sim N(1.01, 0.004)$, we are considering the effect of adding the random variable W to itself. In general, there is no reason why we cannot combine variables $W_1 \sim N(\mu_1, \sigma_1^2)$ and $W_2 \sim N(\mu_2, \sigma_2^2)$. This might happen if the DIY store bought in two similar products from two different manufacturers.

Before we can solve such problems, we need to obtain some results concerning the behaviour of random variables.

Functions of several random variables

Note that we shall quote results only for the continuous case. The results for the discrete case are similar with integration replaced by summation. We will omit the mathematics leading to these results.

Key Point 4

- If $X_1, X_2, \cdots + X_n$ are n random variables then

$$\mathsf{E}(X_1 + X_2 + \cdots + X_n) = \mathsf{E}(X_1) + \mathsf{E}(X_2) + \ldots \mathsf{E}(X_n)$$

- If $X_1, X_2, \ldots X_n$ are n independent random variables then

$$\mathsf{V}(X_1 + X_2 + \cdots + X_n) = \mathsf{V}(X_1) + \mathsf{V}(X_2) + \cdots + \mathsf{V}(X_n)$$

and more generally

$$\mathsf{V}(X_1 \pm X_2 \pm \cdots \pm X_n) = \mathsf{V}(X_1) + \mathsf{V}(X_2) + \cdots + \mathsf{V}(X_n)$$

Example 16

Solve Problem 1 from the previous page. You may assume that the sum and difference of two normal random variables are themselves normal.

Solution

Consider the random variable $C - S$. Using the results above we know that

$$C - S \sim N(3.010 - 3.0, 0.004^2 + 0.003^2) \quad \text{i.e,} \quad C - S \sim N(0.01, 0.005^2)$$

Hence $\quad \mathsf{P}(C - S > 0) = \mathsf{P}(Z > \dfrac{0 - 0.01}{0.005} = -2) = 0.9772$

This result implies that in 2.28% of cases it will be impossible to fit the shaft to the sleeve.

Task

Solve Problem 2 from the previous page. You may assume that the sum and difference of two normal random variables are themselves normal.

Your solution

Answer

If W_{12} is the random variable representing the combined weight of the two boxes then

$$W_{12} \sim N(2.02, 0.008)$$

Hence

$$\text{P}(W_{12} > 2.03) = \text{P}(Z > \frac{2.03 - 2.02}{\sqrt{0.008}} = 0.1118) = 0.5 - 0.0445 = 0.4555$$

The result implies that the customer has about a 46% chance of finding that the weight of the two boxes combined is greater than 2.03 kilograms.

Exercises

1. Batteries of type A have mean voltage 6.0 (volts) and variance 0.0225 (volts^2). Type B batteries have mean voltage 12.0 and variance 0.04. If we form a series connection containing one of each type what is the probability that the combined voltage exceeds 17.4?

2. Nuts and bolts are made separately and paired at random. The nuts' diameters, in mm, are independently $N(10, 0.02)$ and the bolts' diameters, in mm, are independently $N(9.5, 0.02)$. Find the probability that a bolt is too large for its nut.

3. Certain cutting tools have lifetimes, in hours, which are independent and normally distributed with mean 300 and variance 10000.

   (a) Find the probability that

   (i) the total life of three tools is more than 1000 hours.

   (ii) the total life of four tools is more than 1000 hours.

   (b) In a factory each tool is replaced when it fails. Find the probability that exactly four tools are needed to accumulate 1000 hours of use.

   (c) Explain why the first sentence in this question can only be approximately, not exactly, true.

4. A firm produces articles whose length, X, in cm, is normally distributed with nominal mean $\mu = 4$ and variance $\sigma^2 = 0.1$. From time to time a check is made to see whether the value of μ has changed. A sample of ten articles is taken, the lengths are measured, the sample mean length $\bar{X}$ is calculated, and the process is adjusted if $\bar{X}$ lies outside the range (3.9, 4.1). Determine the probability, α, that the process is adjusted as a result of a sample taken when $\mu = 4$. Find the smallest sample size n which would make $\alpha \leq 0.05$.

Answers

1. $X_A \sim N(6, 0.0225)$ $\qquad\qquad X_B \sim N(12, 0.04)$
   Series $X = X_A + X_B \sim N(18, 0.0625)$ as variances always add

$$\mathsf{P}(X > 17.4) = \mathsf{P}\left(Z > \frac{-0.6}{0.25}\right) = 0.5 + \mathsf{P}\left(0 < Z < \frac{0.6}{0.25}\right)$$

$$= 0.5 + \mathsf{P}(0 < Z < 2.4) = 0.5 + 0.4918 = 0.9918$$

2. Let the diameter of a nut be N. Let the diameter of a bolt be B. A bolt is too large for its nut if $N - B < 0$.

$$\begin{aligned}
\mathsf{E}(N-B) &= 10 - 9.5 = 0.5\\
\mathsf{V}(N-B) &= 0.02 + 0.02 = 0.04\\
N - B &\sim N(0.5, 0.04)\\
\mathsf{P}(N-B<0) &= \mathsf{P}\left(\frac{N-B-0.5}{0.2} < \frac{0-0.5}{0.2}\right) = \mathsf{P}(Z < -2.5)\\
&= \Phi(-2.5) = 1 - \Phi(2.5) = 1 - 0.99379\\
&= 0.00621.
\end{aligned}$$

   The probability that a bolt is too large for its nut is 0.00621.

3. (a) Let the lifetime of tool i be T_i.

(i)

$$\begin{aligned}
\mathsf{E}(T_1+T_2+T_3) &= 900\\
\mathsf{V}(T_1+T_2+T_3) &= 30000\\
(T_1+T_2+T_3) &\sim N(900, 30000)\\
\mathsf{P}(T_1+T_2+T_3 > 1000) &= \mathsf{P}\left(\frac{T_1+T_2+T_3-900}{\sqrt{30000}}\right) = \mathsf{P}(Z > 0.57735)\\
&= 1 - \Phi(0.57735) = 1 - 0.7181 = 0.2819
\end{aligned}$$

(ii)

$$\begin{aligned}
\mathsf{E}(T_1+T_2+T_3+T_4) &= 1200\\
\mathsf{V}(T_1+T_2+T_3+T_4) &= 40000\\
(T_1+T_2+T_3+T_4) &\sim N(1200, 40000)\\
\mathsf{P}(T_1+T_2+T_3+T_4 > 1000) &= \mathsf{P}\left(\frac{T_1+T_2+T_3+T_4-1200}{\sqrt{40000}} > \frac{1000-1200}{\sqrt{40000}}\right)\\
&= \mathsf{P}(Z > -1)\\
&= 1 - \Phi(-1) = \Phi(1) = 0.8413
\end{aligned}$$

Answers

(b) Let the number of tools needed be N.

$$\begin{aligned} \mathsf{P}(N \leq 3) &= \mathsf{P}(N=1)+\mathsf{P}(N=2)+\mathsf{P}(N=3) \\ &= \mathsf{P}(T_1+T_2+T_3>1000) = 0.2819 \\ \mathsf{P}(N \leq 4) &= \mathsf{P}(N=1)+\mathsf{P}(N=2)+\mathsf{P}(N=3)+\mathsf{P}(N=4) \\ &= \mathsf{P}(T_1+T_2+T_3+T_4>1000) = 0.8413. \end{aligned}$$

Hence $\mathsf{P}(N=4) = \mathsf{P}(N \leq 4) - \mathsf{P}(N \leq 3) = 0.8413 - 0.2819 = 0.5594.$

(c) Lifetimes can not be negative. The normal distribution assigns non-zero probability density to negative values so it can only be an approximation in this case.

4.

$$\begin{aligned} X &\sim N(4, 0.1) \\ X_1 + \cdots + X_{10} &\sim N(40, 1) \\ \bar{X} = (X_1 + \cdots + X_{10})/10 &\sim N(4, 0.01) \end{aligned}$$

By symmetry $\mathsf{P}(\bar{X} < 3.9) = \mathsf{P}(\bar{X} > 4.1)$.

$$\begin{aligned} \mathsf{P}(\bar{X} < 3.9) &= \mathsf{P}\left(\frac{\bar{X}-4}{0.1} < \frac{3.9-4}{0.1}\right) = \mathsf{P}(Z < -1) \\ &= \Phi(-1) = 1 - \Phi(1) = 1 - 0.8413 \\ &= 0.1587 \end{aligned}$$

More generally

$$\begin{aligned} X_1 + \cdots + X_n &\sim N(4n, 0.1n) \\ \bar{X} = (X_1 + \cdots + X_n)/n &\sim N(4, 0.1/n) \\ \mathsf{P}(\bar{X} < 3.9) &= \mathsf{P}\left(\frac{\bar{X}-4}{\sqrt{0.1/n}} < \frac{3.9-4}{\sqrt{0.1/n}}\right) = \mathsf{P}\left(Z < \frac{-0.1}{\sqrt{0.1/n}} = -\sqrt{0.1n}\right) \\ &= \Phi(-\sqrt{0.1n}) = 1 - \Phi(\sqrt{0.1n}) \\ \alpha &= 2[1 - \Phi(\sqrt{0.1n})] \end{aligned}$$

We require $\alpha \leq 0.05$.

$$\begin{aligned} 2[1 - \Phi(\sqrt{0.1n})] \leq 0.05 &\Leftrightarrow 1 - \Phi(\sqrt{0.1n}) \leq 0.025 \\ &\Leftrightarrow \Phi(\sqrt{0.1n}) \geq 0.975 \\ &\Leftrightarrow \sqrt{0.1n} \geq 1.96 \\ &\Leftrightarrow n \geq 10 \times 1.96^2 = 38.416 \end{aligned}$$

The smallest sample size which satisfies this is $n = 39$.

Table 1: The Standard Normal Probability Integral

$Z=\frac{x-\mu}{\sigma}$	0	1	2	3	4	5	6	7	8	9
0	0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
.1	0398	0438	0478	0517	0577	0596	0636	0675	0714	0753
.2	0793	0832	0871	0909	0948	0987	1026	1064	1103	1141
.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
.4	1555	1591	1628	1664	1700	1736	1772	1808	1844	1879
.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
.6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
.7	2580	2611	2642	2673	2703	2734	2764	2794	2822	2852
.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4865	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4946	4947	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2.9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986

	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9
	4987	4990	4993	4995	4997	4998	4998	4999	4999	4999