

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 ...

- ① be familiar with the results and concepts met in the study of probability
- ② be familiar with the normal distribution



Learning Outcomes

After completing this Section you should be able to ...

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

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 3cm 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 similar products from 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

- If X_1, X_2, \dots, X_n are n random variables then
$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$
- If X_1, X_2, \dots, X_n are n independent random variables then
$$V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n)$$

and more generally

$$V(X_1 \pm X_2 \pm \dots \pm X_n) = V(X_1) + V(X_2) + \dots + V(X_n)$$

Example *Solve Problem 1* 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

$$P(C - S > 0) = P\left(Z > \frac{0 - 0.01}{0.005} = -2\right) = 0.9772$$

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



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

Your solution

If W_2 is the random variable representing the combined weight of the two boxes then

$$W_2 \sim N(2.02, 0.008)$$

Hence

$$P(W_2 > 2.03) = P\left(Z > \frac{2.03 - 2}{\sqrt{0.008}}\right) = 0.3354 = 0.5 - 0.1312 = 0.3688$$

The result implies that the customer has about a 37% 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 and variance 0.0225 (measurements in volts). 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. Find the probability that
 - (a) the total life of three tools is more than 1000 hours.
 - (b) the total life of four tools is more than 1000 hours.

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.

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

4. A firm produces articles the length, X , in cm., of which 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 $\alpha \leq 0.05$.

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

$$P(X > 17.4) = P\left(Z > \frac{-0.6}{0.25}\right) = 0.5 + P\left(0 < Z < \frac{0.6}{0.25}\right) = 0.5 + P(0 < Z < 2.4) = 0.5 + 0.4918 = 0.9918$$

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 $N - B > 0$.

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

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

3. Let the lifetime of tool i be T_i .

(a)

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

(b)

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

Continued Let the number of tools needed be N .

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

Hence $\Pr(N = 4) = \Pr(N \leq 4) - \Pr(N \leq 3) = 0.8413 - 0.2819 = 0.5594$.

Lifetimes can not be negative. The normal distribution assigns nonzero probability density to negative values so it can only be an approximation in this case.

4.

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

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

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

More generally

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

We require $\alpha \leq 0.05$.

$$2[1 - \Phi(\sqrt{0.1n})] \leq 0.05 \Leftrightarrow 1 - \Phi(\sqrt{0.1n}) \leq 0.025$$

$$\Leftrightarrow \Phi(\sqrt{0.1n}) \geq 0.975$$

$$\Leftrightarrow \sqrt{0.1n} \geq 1.96$$

$$\Leftrightarrow n \geq 10 \times 1.96^2 = 38.416$$

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