

13.5 Asking Awkward Questions

WORKSHEET

The question

- (i) X_1 and X_2 are both random variables which take values $x_1, x_2, \dots, x_n$, with probabilities $a_1, a_2, \dots, a_n$ and $b_1, b_2, \dots, b_n$ respectively.

The value of random variable Y is defined to be that of X_1 with probability p and that of X_2 with probability $q = 1 - p$.

If X_1 has mean μ_1 and variance σ_1^2 , and X_2 has mean μ_2 and variance σ_2^2 , find the mean of Y and show that the variance of Y is $p\sigma_1^2 + q\sigma_2^2 + pq(\mu_1 - \mu_2)^2$.

- (ii) To find the value of random variable B , a fair coin is tossed and a fair six-sided die is rolled. If the coin shows heads, then $B = 1$ if the die shows a six and $B = 0$ otherwise; if the coin shows tails, then $B = 1$ if the die does not show a six and $B = 0$ if it does.

The value of Z_1 is the sum of n independent values of B , where n is large. Show that Z_1 is a Binomial random variable with probability of success $\frac{1}{2}$. Using a Normal approximation, show that the probability that Z_1 is within 10% of its mean tends to 1 as $n \rightarrow \infty$.

- (iii) To find the value of random variable Z_2 , a fair coin is tossed and n fair six-sided dice are rolled, where n is large. If the coin shows heads, then the value of Z_2 is the number of dice showing a six; if the coin shows tails, then the value of Z_2 is the number of dice not showing a six.

Use part (i) to write down the mean and variance of Z_2 . Explain why a Normal distribution with this mean and variance will not be a good approximation to the distribution of Z_2 . Show that the probability that Z_2 is within 10% of its mean tends to 0 as $n \rightarrow \infty$.

Hints

Try the question on your own first. If you get stuck, read the hints *one at a time*, and go back to the question after each one. Hints marked *Pause and try* are the key steps.



Ask yourself Y is X_1 with probability p and X_2 with probability q . What is $E(Y)$? What is $E(Y^2)$?

STAGE 2 Independent randomisation



Ask yourself What is $P(B = 1)$? Are the n values of B independent?

STAGE 3 One coin for everybody



Ask yourself Now one coin toss decides for all n dice. Given heads, what is the distribution of Z_2 ? Given tails? Use part (i).



Ask yourself Where is most of the probability? Is there much near the mean $\frac{n}{2}$?

Try this next

- The twins.** (From STEP II, 2005.) Twins share a computer; the one online is Anna with probability p , who answers truthfully with probability a , or Bella, who does so with probability b . I ask them to toss a coin and say whether it came down heads, and get “yes”. Show that the probability that it was heads is $\frac{1}{2}$ if and only if $2(ap + bq) = 1$. If I then ask again and get “no”, show that the probability of heads is $\frac{1}{2}$; if I get “yes” again, show that it is $\frac{1}{2}$ provided $2(ap + bq) = 1$.
- Estimating π .** In a Warner survey with $p = \frac{3}{4}$, 400 people answer and 160 say yes. Estimate the proportion π in group A , and find the approximate standard error of your estimate.
- Total variance.** Prove the law of total variance for a discrete hidden variable C with several values.
- Shared coins.** In part (iii), what is the probability that $Z_2 > \frac{n}{2}$? Explain why, for large n , it is close to $\frac{1}{2}$ even though Z_2 is almost never close to $\frac{n}{2}$.

Further reading

- Stanley Warner, “Randomized response: a survey technique for eliminating evasive answer bias”, *Journal of the American Statistical Association* 60 (1965).

- ▶ **Cynthia Dwork & Aaron Roth, *The Algorithmic Foundations of Differential Privacy*** (free online). The modern theory; the first chapters are accessible.
- ▶ **David Spiegelhalter, *The Art of Statistics*** (Pelican), on surveys and their pitfalls.

The full guided solution, and what the question leads to at university, is in Section 13.5:
https://keithmcnulty.github.io/step_book_mcnulty/ch13/awkward-questions.html