Question (1991 STEP I Q15)

A fair coin is thrown n times. On each throw, 1 point is scored for a head and 1 point is lost for a tail. Let S_n be the points total for the series of n throws, i.e. $S_n = X_1 + X_2 + \cdots + X_n$, where

$$X_j = \begin{cases} 1 & \text{if the } j \text{ th throw is a head} \\ -1 & \text{if the } j \text{ th throw is a tail.} \end{cases}$$

(i) If $n = 10\,000$, find an approximate value for the probability that $S_n > 100$.

(ii) Find an approximate value for the least n for which $\mathrm{P}(S_n > 0.01n) < 0,01$.

Suppose that instead no points are scored for the first throw, but that on each successive throw, 2 points are scored if both it and the first throw are heads, two points are deducted if both are tails, and no points are scored or lost if the throws differ. Let Y_k be the score on the kth throw, where $2 \leqslant k \leqslant n$. Show that $Y_k = X_1 + X_k$. Calculate the mean and variance of each Y_k and determine whether it is true that

$$\mathrm{P}(Y_2 + Y_3 + \cdots + Y_n > 0.01(n-1)) \to 0 \qquad \text{as } n \to \infty.$$

Notice that $\mathbb{E}(X_i) = 0, \mathbb{E}(X_i^2) = 1$ and so $\mathbb{E}(S_n) = 0, \mathrm{Var}(S_n) = n$.

(i) Then by the central limit theorem (or alternatively the normal approximation to the binomial),

$$\begin{aligned} \mathbb{P}(S_n > 100) &\underbrace{\approx}_{\text{CLT}} \mathbb{P}\left(Z > \frac{100}{\sqrt{10\,000}}\right) \\ &= \mathbb{P}(Z > 1) \\ &= 1 - \Phi(1) \\ &\approx 15.9\% \end{aligned}$$

(ii)

$$\begin{aligned} \mathbb{P}(S_n > 0.01n) &\approx \mathbb{P}\left(Z > \frac{0.01n}{\sqrt{n}}\right) \\ &= \mathbb{P}(Z > 0.01\sqrt{n}) \\ &= 1 - \Phi(0.01\sqrt{n}) \\ &< 0.01 \\ \Phi^{-1}(0.01) &= -2.3263\ldots \\ \Rightarrow \quad 0.01\sqrt{n} &= 2.3263\ldots \\ \Rightarrow \quad n &\approx 233^2 \end{aligned}$$

1st throw	kth throw	$X_1 + X_k$	Y_k
head	head	$1+1$	2
head	tail	$1-1$	0
tail	head	$-1+1$	0
tail	tail	$-1-1$	-2

Across all possible cases $Y_k = X_1 + X_k$ so therefore these random variables are equal.

$$\begin{aligned}\mathbb{E}(Y_k) &= \mathbb{E}(X_1) + \mathbb{E}(Y_k) \\ &= 0 + 0 = 0\end{aligned}$$

$$\begin{aligned}\text{Var}(Y_k) &= \text{Var}(X_1) + \text{Var}(X_k) \\ &= 2\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left(\sum_{k=2}^{n} Y_k\right) &= 0 \\ \text{Var}\left(\sum_{k=2}^{n} Y_k\right) &= 2(n-1)\end{aligned}$$

Therefore approximately $\displaystyle\sum_{k=2}^{n} Y_k \approx N(0, 2(n-1))$

$$\begin{aligned}\mathbb{P}\left(\sum_{k=2}^{n} Y_k > 0.01(n-1)\right) &\approx \mathbb{P}\left(Z > \frac{0.01(n-1)}{\sqrt{2(n-1)}}\right) \\ &= \mathbb{P}\left(Z > c\sqrt{n-1}\right) \\ &\to 0 \text{ as } n \to \infty\end{aligned}$$

Question (1992 STEP I Q14)

The average number of pedestrians killed annually in road accidents in Poldavia during the period 1974-1989 was 1080 and the average number killed annually in commercial flight accidents during the same period was 180. Discuss the following newspaper headlines which appeared in 1991. (The percentage figures in square brackets give a rough indication of the weight of marks attached to each discussion.)

(i) [10%] Six Times Safer To Fly Than To Walk. 1974-1989 Figures Prove It.

(ii) [10%] Our Skies Are Safer. Only 125 People Killed In Air Accidents In 1990.

(iii) [30%] Road Carnage Increasing. 7 People Killed On Tuesday.

(iv) [50%] Alarming Rise In Pedestrian Casualties. 1350 Pedestrians Killed In Road Accidents During 1990.

Question (1996 STEP II Q14)

The random variable X is uniformly distributed on $[0, 1]$. A new random variable Y is defined by the rule

$$Y = \begin{cases} 1/4 & \text{if } X \leqslant 1/4, \\ X & \text{if } 1/4 \leqslant X \leqslant 3/4 \\ 3/4 & \text{if } X \geqslant 3/4. \end{cases}$$

Find $\mathrm{E}(Y^n)$ for all integers $n \geqslant 1$. Show that $\mathrm{E}(Y) = \mathrm{E}(X)$ and that

$$\mathrm{E}(X^2) - \mathrm{E}(Y^2) = \frac{1}{24}.$$

By using the fact that $4^n = (3+1)^n$, or otherwise, show that $\mathrm{E}(X^n) > \mathrm{E}(Y^n)$ for $n \geqslant 2$. Suppose that Y_1, Y_2, ...are independent random variables each having the same distribution as Y. Find, to a good approximation, K such that

$$\mathrm{P}(Y_1 + Y_2 + \cdots + Y_{240000} < K) = 3/4.$$

None

Question (1997 STEP I Q12)

An experiment produces a random number T uniformly distributed on $[0, 1]$. Let X be the larger root of the equation

$$x^2 + 2x + T = 0.$$

What is the probability that $X > -1/3$? Find $\mathrm{E}(X)$ and show that $\mathrm{Var}(X) = 1/18$. The experiment is repeated independently 800 times generating the larger roots X_1, X_2, ..., X_{800}. If

$$Y = X_1 + X_2 + \cdots + X_{800}.$$

find an approximate value for K such that

$$\mathrm{P}(Y \leqslant K) = 0.08.$$

Question (1998 STEP III Q14)

A hostile naval power possesses a large, unknown number N of submarines. Interception of radio signals yields a small number n of their identification numbers X_i $(i = 1, 2, ..., n)$, which are taken to be independent and uniformly distributed over the continuous range from 0 to N. Show that Z_1 and Z_2, defined by

$$Z_1 = \frac{n+1}{n} \max\{X_1, X_2, ..., X_n\} \qquad \text{and} \qquad Z_2 = \frac{2}{n} \sum_{i=1}^{n} X_i \ ,$$

both have means equal to N. Calculate the variance of Z_1 and of Z_2. Which estimator do you prefer, and why?

Question (2000 STEP II Q14)

The random variables X_1, X_2, ... , X_{2n+1} are independently and uniformly distributed on the interval $0 \leq x \leq 1$. The random variable Y is defined to be the median of X_1, X_2, ... , X_{2n+1}. Given that the probability density function of Y is $g(y)$, where

$$\mathrm{g}(y) = \begin{cases} ky^n(1-y)^n & \text{if } 0 \leqslant y \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

use the result

$$\int_0^1 y^r(1-y)^s \,\mathrm{d}y = \frac{r!s!}{(r+s+1)!}$$

to show that $k = (2n+1)!/(n!)^2$, and evaluate $\mathbb{E}(Y)$ and $\mathrm{V}ar\,(Y)$. Hence show that, for any given positive number d, the inequality

$$\mathbb{P}\left(|Y - 1/2| < d/\sqrt{n}\right) < \mathbb{P}\left(|\bar{X} - 1/2| < d/\sqrt{n}\right)$$

holds provided n is large enough, where $\bar{X}$ is the mean of X_1, X_2, ... , X_{2n+1}. [You may assume that Y and $\bar{X}$ are normally distributed for large n.]

Question (2001 STEP II Q14)

Two coins A and B are tossed together. A has probability p of showing a head, and B has probability $2p$, independent of A, of showing a head, where $0 < p < \frac{1}{2}$. The random variable X takes the value 1 if A shows a head and it takes the value 0 if A shows a tail. The random variable Y takes the value 1 if B shows a head and it takes the value 0 if B shows a tail. The random variable T is defined by

$$T = \lambda X + \tfrac{1}{2}(1-\lambda)Y.$$

Show that $\mathbb{E}(T) = p$ and find an expression for $\mathrm{Var}(T)$ in terms of p and λ. Show that as λ varies, the minimum of $\mathrm{Var}(T)$ occurs when

$$\lambda = \frac{1-2p}{3-4p}\,.$$

The two coins are tossed n times, where $n > 30$, and $\overline{T}$ is the mean value of T. Let b be a fixed positive number. Show that the maximum value of $\mathbb{P}\big(|\overline{T} - p| < b\big)$ as λ varies is approximately $2\Phi(b/s) - 1$, where Φ is the cumulative distribution function of a standard normal variate and

$$s^2 = \frac{p(1-p)(1-2p)}{(3-4p)n}\,.$$

None

Question (2001 STEP III Q14)

A random variable X is distributed uniformly on $[\,0\,,\,a\,]$. Show that the variance of X is $\frac{1}{12}a^2$. A sample, X_1 and X_2, of two independent values of the random variable is drawn, and the variance V of the sample is determined. Show that $V = \frac{1}{4}\left(X_1 - X_2\right)^2$, and hence prove that $2V$ is an unbiased estimator of the variance of X. Find an exact expression for the probability that the value of V is less than $\frac{1}{12}a^2$ and estimate the value of this probability correct to one significant figure.

Question (2009 STEP III Q13) **(i)** The point P lies on the circumference of a circle of unit radius and centre O. The angle, θ, between OP and the positive x-axis is a random variable, uniformly distributed on the interval $0 \le \theta < 2\pi$. The cartesian coordinates of P with respect to O are (X, Y). Find the probability density function for X, and calculate $\text{Var}(X)$. Show that X and Y are uncorrelated and discuss briefly whether they are independent.

(ii) The points P_i $(i = 1, 2, \ldots, n)$ are chosen independently on the circumference of the circle, as in part (i), and have cartesian coordinates (X_i, Y_i). The point $\overline{P}$ has coordinates $(\overline{X}, \overline{Y})$, where $\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$ and $\overline{Y} = \dfrac{1}{n}\sum\limits_{i=1}^{n} Y_i$. Show that $\overline{X}$ and $\overline{Y}$ are uncorrelated. Show that, for large n, $\mathbb{P}\left(|\overline{X}| \le \sqrt{\dfrac{2}{n}}\right) \approx 0.95\,.$