Coursework 6

1. Prove that, if $X_1, X_2, ..., X_n$ is a sequence of independent random variables with $E(X_i) = \mu_j$, $\mathrm{Var}(X_j) = \sigma_j^2$ then

$$\mathrm{Var}\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} \sigma_j^2.$$

Solution: By the definition of variance,

$$\begin{aligned}
\mathrm{Var}\left(\sum_{j=1}^{n} X_j\right) &= E\left(\sum_{j=1}^{n} X_j - E\left(\sum_{j=1}^{n} X_j\right)\right)^2 \\
&= E\left(\sum_{j=1}^{n} X_j - \sum_{j=1}^{n} E[X_j]\right)^2 \\
&= E\left(\sum_{j=1}^{n} (X_j - E(X_j))\right)^2 \\
&= E\left(\sum_{j=1}^{n} (X_j - \mu_j)\right)^2 \\
&\overset{(*)}{=} \sum_{j=1}^{n} E(X_j - \mu_j)^2 + 2 \sum_{1 \le j < i \le n} E\left((X_j - \mu_j)(X_i - \mu_i)\right) \\
&\overset{(**)}{=} \sum_{j=1}^{n} \mathrm{Var}(X_j) + 2 \sum_{1 \le j < i \le n} \mathrm{Cov}(X_j, X_i) \\
&\overset{(***)}{=} \sum_{j=1}^{n} \mathrm{Var}(X_j) = \sum_{j=1}^{n} \sigma_j^2.
\end{aligned}$$

where
$(*)$ is due to the rule $(\sum_{j=1}^{n} a_j)^2 = \sum_{j=1}^{n} a_j^2 + 2 \sum_{1 \le j < i \le n} a_j a_i$,

$(**)$ is due to the definition of covariance
$(***)$ holds because X_i and X_j are independent and therefore

$$E\left(\left(X_j-\mu_j\right)\left(X_i-\mu_i\right)\right)=E\left(X_j-\mu_j\right)\times E\left(X_i-\mu_i\right)=0.$$

2. Suppose that I roll a (fair) die repeatedly. Let S_n be the total number of 5's or 6's that I observe after throwing the die n times. What is the

$$\lim_{n\to\infty} P(0.3n < S_n < 0.4n)?$$

Solution: We use the law of large numbers.

Let X_k be the random variable that is 1 if the roll is a 5 or 6 and zero otherwise. Then $S_n=\sum_{k=1}^n X_k$. We are interested in the event

$$\{0.3n < S_n < 0.4n\}=\{0.3<\frac{S_n}{n}<0.4\}=\{0.3<Y_n<0.4\}$$

where $Y_n=S_n/n$ as in the law of large numbers. We see that $E(X_k)=1/3$ and that $\mathrm{Var}(X_k)=2/9<\infty$.

Let $\varepsilon=0.01$.

$$\{0.3<Y_n<0.4\}\supset\{|Y_n-1/3|<0.03\}$$

Hence

$$P(0.3n<S_n<0.4n)=P(0.3<Y_n<0.4)>P(|Y_n-1/3|<0.03)$$

which tends to 1 by the Law of large numbers. Hence

$$P(0.3n<S_n<0.4n)\to 1.$$

Note you might be inclined to pick $\varepsilon=1/3-0.3=1/30$: you can if you like but since you can pick any ε pick something easy!

3. $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean μ and variance σ^2. How large a sample must be taken in order that you can be 95% certain that the sample mean

$$\bar{X}_n = \frac{1}{n} \sum_{n}^{i=1} X_i \quad \text{is within } 0.1\sigma \text{ of } \mu?$$

Solution: First of all, $\bar{X}_n$ has mean μ and variance σ^2/n (prove that). We have to find n such that

$P(|\bar{X}_n - \mu| \leq 0.1\sigma) \geq 0.95$ or, equivalently, $P(|\bar{X}_n - \mu| > 0.1\sigma) \leq 0.05$.

By Chebyshev's inequality, we have

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2/n}{\varepsilon^2}.$$

In our case, $\varepsilon = 0.1\sigma$ and

$$P\left(|\bar{X}_n - \mu| \geq 0.1\sigma\right) \leq \frac{\sigma^2}{n(0.1\sigma)^2} = \frac{100}{n}.$$

It remains to find n such that $\frac{100}{n} \leq 0.05$. Hence $n \geq \frac{100}{0.05} = 2000$. In words, if n is 2000 or larger, then $\bar{X}_n$ is within 0.1σ of μ.

4. Suppose that I measure the heights of 100 people in London. A person's height has mean 160cm and standard deviation 15cm. Find the (approximate) probability that the mean height of these 100 people I measure is over 163cm. Assume each person's height is independent from the others'. Express your answer in terms of the Φ function.

   Solution: Let X_i be the height in cm of the ith person I measure. The X_i are $n = 100$ independent random variables with mean $\mu = 160$ and variance $\sigma^2 = 15^2$. We have to estimate the probability

$$
\begin{aligned}
P\left(\frac{1}{100}\sum_{j=1}^{100} X_j \geq 163\right) &= P\left(\sum_{j=1}^{100} X_j \geq 100\cdot 163\right) \\
&= P\left(\frac{\sum_{j=1}^{100} X_j - n\mu}{\sqrt{n}\sigma} \geq \frac{100\cdot 163 - n\mu}{\sqrt{n}\sigma}\right) \\
&= P\left(\frac{\sum_{j=1}^{100} X_j - n\mu}{\sqrt{n}\sigma} \geq \frac{100\cdot 163 - 100\cdot 160}{10\cdot 15}\right) \\
&= P\left(\frac{\sum_{j=1}^{100} X_j - n\mu}{\sqrt{n}\sigma} \geq 2\right) = 1 - \Phi(2) = 0.023.
\end{aligned}
$$

Explanation. We have transformed both sides of the initial inequality so that to turn the left hand side into the expression which appears in the CLT. Namely, by the CLT the random variable $Z_n := \frac{\sum_{j=1}^{n} X_j - n\mu}{\sqrt{n}\sigma}$ is approximately normal, $N(0,1)$. Thus the probability

$$P(Z_n > x) = 1 - P(Z_n \leq x) =\approx 1 - \Phi(x).$$

In our case $x = 2$ and we use the normal distribution table to see that this probability is approximately 0.977.

5. [Gambler's ruin problem] Suppose that we are gambling repetitively on a game with probability of losing £1 in each gamble 0.55 and winning £1 with probability 0.45. Starting from an initial capital of £20. Show that the probability we have not gone bankrupt after 1000 games is (approximately) at most 0.0057.

Solution: The gambler just plays the game for 1000 times and looks at the distribution of the amount of money she has.

Let X_k be the random variable that is 1 if she wins the k-th game and -1 if she loses it.

Then the capital at time 1000 is

$$C_{1000} = \pounds\left(20 + \sum_{k=1}^{1000} X_k\right).$$

Now each X_k has

$$E[X_k] = -0.1$$

and

$$\mathrm{Var(X_k)} = 0.99.$$

Hence by the independence of the X_i's, we have

$$S_{1000} = \sum_{k=1}^{1000} X_k$$

has mean -100 and variance 990.

By the approximate Central Limit Theorem

$$Z_{1000} = \frac{S_{1000} - (-100)}{\sqrt{990}} \sim N(0,1).$$

Note also that the capital (in pounds) at time 1000 is approximately

$$C_{1000} = 20 + \sum_{k=1}^{1000} X_k = 20 + S_{1000}.$$

Hence, using the above distribution, we can estimate the probability that the gambler's capital is negative at time 1000, by calculating

$$\begin{aligned}
P(C_{1000} < 0) &= P(S_{1000} < -20) \\
&= P\Big(\frac{S_{1000} - (-100)}{\sqrt{990}} < \frac{-20 - (-100)}{\sqrt{990}}\Big) \\
&= P(Z_{1000} < 2.54) \\
&\approx \int_{-\infty}^{2.54} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \Phi(2.54) = 0.9943.
\end{aligned}$$

since Z_{1000} is (approximately) a standard Normal random variable.

Now the event that we go bankrupt, at any time between 1 and 1000, is a subset of the event we have negative money at time 1000. Hence,

$$P(\text{Go bankrupt}) \geq P(C_{1000} < 0) = 0.9943$$

and

$$P(\text{Not gone bankrupt}) = 1 - P(\text{Go bankrupt}) \leq 1 - 0.9943 = 0.0057.$$

Therefore, the probability that we have not gone bankrupt is at most 0.57%.