PROBABILITY

MOISE GABRIEL

PROBLEM SHEET 8

1. (a) $f_{X,Y}(x,y) = c_1\left(x^2 + \frac{1}{3}xy\right)$, $x \in (0,1)$, $y \in (0,2)$

   (b) $f_{X,Y}(x,y) = c_2\, e^{-x-y}$, $0 < x < y < \infty$

(a) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx = 1 \iff \int_0^1\int_0^2 c_1\left(x^2+\frac{1}{3}xy\right)dy\,dx = 1 \iff$

$\iff \int_0^1 \left[c_1x^2y + c_1x\frac{y^2}{6}\right]_0^2 dx = 1 \iff \int_0^1\left(2c_1x^2 + \frac{2}{3}c_1x\right)dx = 1 \iff$

$\iff \left[\frac{2}{3}c_1x^3 + \frac{1}{3}c_1x^2\right]_0^1 = 1 \iff \frac{2}{3}c_1 + \frac{1}{3}c_1 = 1 \iff \boxed{c_1 = 1}$

(b) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx = 1 \iff \int_0^{\infty}\int_x^{\infty} c_2e^{-x-y}\,dy\,dx = 1 \iff$

$\iff \int_0^{\infty}\left[c_2e^{-x}\cdot(-e^{-y})\right]_x^{\infty} dx = 1 \iff \int_0^{\infty} c_2e^{-2x}dx = 1 \iff \left[-\frac{1}{2}c_2\,e^{-2x}\right]_0^{\infty} = 1 \iff$

$\iff \frac{1}{2}c_2 = 1 \iff \boxed{c_2 = 2}$

- are X and Y independent?

(a) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^2\left(x^2+\frac{1}{3}xy\right)dy = \left[x^2y + \frac{1}{6}xy^2\right]_0^2 = 2x^2 + \frac{2}{3}x$

$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^1\left(x^2+\frac{1}{3}xy\right)dx = \left[\frac{1}{3}x^3 + \frac{1}{6}x^2y\right]_0^1 = \frac{1}{3} + \frac{1}{6}y$

$f_X(x)f_Y(y) = \left(2x^2+\frac{2}{3}x\right)\left(\frac{1}{3}+\frac{1}{6}y\right) = \frac{2}{3}x^2 + \frac{2}{9}x + \frac{1}{3}x^2y + \frac{1}{9}xy \neq f_{X,Y}(x,y) \Rightarrow$

$\Rightarrow$ X and Y are NOT independent

(b) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_x^{\infty} 2e^{-x-y}\,dy = \left[-2e^{-x-y}\right]_x^{\infty} = 2e^{-2x}$

$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^{\infty} 2e^{-x-y}\,dy = \left[-2e^{-x-y}\right]_0^{\infty} = 2e^{-y}$

$f_X(x)\,f_Y(y) = 2e^{-2x}\cdot 2e^{-y} = 4e^{-2x-y} \neq f_{X,Y}(x,y) \Rightarrow$ X and Y are NOT independent

- find the marginal probability density functions of X and of Y

(a) $f_X(x) = 2x^2 + \frac{2}{3}x$ and $f_Y(y) = \frac{1}{3} + \frac{1}{6}y$ (from above)

(b) $f_X(x) = 2e^{-2x}$ and $f_Y(y) = 2e^{-y}$ (from above)

- find $\mathbb{P}\left(X \le \frac{1}{2}, Y \le 1\right)$

(a) $\mathbb{P}\left(X \le \frac{1}{2}, Y \le 1\right) = \int_{-\infty}^{\frac{1}{2}}\int_{-\infty}^{1} f_{X,Y}(x,y)\,dy\,dx = \int_0^{\frac{1}{2}}\int_0^1\left(x^2 + \frac{1}{3}xy\right)dy\,dx =$

$= \int_0^{\frac{1}{2}}\left[x^2y + \frac{1}{6}xy^2\right]_0^1 dx = \int_0^{\frac{1}{2}}\left(x^2 + \frac{1}{6}x\right)dx = \left[\frac{1}{3}x^3 + \frac{1}{12}x^2\right]_0^{\frac{1}{2}} = \frac{1}{24} + \frac{1}{48} = \frac{3}{48} = \frac{1}{16}.$

(b) $\mathbb{P}\left(X \le \frac{1}{2}, Y \le 1\right) = \int_{-\infty}^{\frac{1}{2}}\int_{-\infty}^{1} f_{X,Y}(x,y)\,dy\,dx = \int_0^{\frac{1}{2}}\int_{\frac{1}{2}}^{1} 2e^{-x-y}\,dy\,dx = \int_0^{\frac{1}{2}}\left[-2e^{-x-y}\right]_{\frac{1}{2}}^{1} dx =$

$$= \int_0^{\frac{1}{2}} \left(2e^{-x-\frac{1}{2}} - 2e^{-x-1}\right) dx = \left[-2e^{-x-\frac{1}{2}} + 2e^{-x-1}\right]_0^{\frac{1}{2}} = 2e^{-\frac{3}{2}} - 4e^{-1} + 2e^{-\frac{1}{2}}.$$

In case (b), if we had $0 < x, y < \infty$ as the region, then:

$$f_{X,Y}(x,y) = C_3\, e^{-x-y}, \quad 0 < x, y < \infty$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx = 1 \iff \int_0^{\infty}\int_0^{\infty} C_3 e^{-x-y}\,dy\,dx = 1 \iff \int_0^{\infty}\left[-C_3 e^{-x-y}\right]_0^{\infty} dx = 1 \iff$$

$$\iff \int_0^{\infty} C_3 e^{-x}\,dx = 1 \iff \left[-C_3 e^{-x}\right]_0^{\infty} = 1 \iff \boxed{C_3 = 1}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^{\infty} e^{-x-y}\,dy = \left[-e^{-x-y}\right]_0^{\infty} = e^{-x}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^{\infty} e^{-x-y}\,dx = \left[-e^{-x-y}\right]_0^{\infty} = e^{-y}$$

$f_X(x)\, f_Y(y) = e^{-x}\cdot e^{-y} = e^{-x-y} = f_{X,Y}(x,y) \Rightarrow$ X and Y are now independent.

2. T_i denotes the time (in hours) at which we see a creature of type i, for $1 \le i \le n$. We suppose that $T_1, T_2, \ldots, T_m$ are independent and that $T_i \sim Exp(\lambda_i)$, $i \in \{1, 2, \ldots, m\}$.

(a) Let $X = \min\{T_1, T_2, \ldots, T_m\}$ be the time at which we see the first creature.

Firstly, we know that:

$f_{T_i}(x) = \lambda_i e^{-\lambda_i x}$, $x \geq 0$ and $F_{T_i}(x) = 1 - e^{-\lambda_i x}$, $x > 0$ for all $1 \le i \le n$.

We then have:

$$\mathbb{P}(X > t) = \mathbb{P}\left(\min\{T_1, T_2, \ldots, T_m\} > t\right) = \mathbb{P}\left(\{T_1 > t\} \cap \{T_2 > t\} \cap \ldots \cap \{T_m > t\}\right) \overset{\text{independence}}{=} \mathbb{P}(T_1 > t)\cdot\mathbb{P}(T_2 > t)\cdot\ldots$$

$$\cdot\,\mathbb{P}(T_m > t) = \left(1 - \mathbb{P}(T_1 \le t)\right)\cdot\left(1 - \mathbb{P}(T_2 \le t)\right)\cdot\ldots\cdot\left(1 - \mathbb{P}(T_n \le t)\right) = e^{-\lambda_1 t}\cdot e^{-\lambda_2 t}\cdot\ldots\cdot e^{-\lambda_m t} \Rightarrow$$

$$\Rightarrow \mathbb{P}(X > t) = e^{-(\lambda_1 + \lambda_2 + \ldots + \lambda_n)t} \Rightarrow \mathbb{P}(X \le t) = 1 - \mathbb{P}(X > t) = 1 - e^{-(\lambda_1 + \lambda_2 + \ldots + \lambda_n)t} \Rightarrow$$

$$\Rightarrow F_X(t) = 1 - e^{-(\lambda_1 + \lambda_2 + \ldots + \lambda_n)t} \Rightarrow \boxed{X \sim Exp(\lambda_1 + \lambda_2 + \ldots + \lambda_n)}$$

(b) For each r.v. T_i we choose the indicator I_i:

$$I_i = \begin{cases} 0, & \text{if we have met the } i^{th} \text{ type of creature by time } 1 \\ 1, & \text{otherwise} \end{cases}$$

$$\mathbb{P}(I_i = 0) = \mathbb{P}(T_i \le 1) = F_{T_i}(1) = 1 - e^{-\lambda_i}$$

$$\mathbb{P}(I_i = 1) = \mathbb{P}(T_i > 1) = 1 - \mathbb{P}(T_i \le 1) = 1 - \mathbb{P}(I_i = 0) = 1 - 1 + e^{-\lambda_i} = e^{-\lambda_i}$$

We deduce that $I_i \sim Ber(e^{-\lambda_i})$ and let A denote the number of creatures we have not met by time 1.

Then, $A = I_1 + I_2 + \ldots + I_m \Rightarrow \mathbb{E}(A) = \mathbb{E}(I_1 + I_2 + \ldots + I_m) = \mathbb{E}(I_1) + \mathbb{E}(I_2) + \ldots + \mathbb{E}(I_m)$

$\mathbb{E}(I_i) = 0\cdot(1 - e^{-\lambda_i}) + 1\cdot e^{-\lambda_i} = e^{-\lambda_i}$ $\Rightarrow$

$$\Rightarrow \boxed{\mathbb{E}(A) = \sum_{i=1}^{n} e^{-\lambda_i}}$$

(c) Let $M = \max\{T_1, T_2, \dots, T_n\}$ be the time until we have met all n different types of creature. Now, we suppose that $\lambda_1 = \lambda_2 = \dots = \lambda_n = 1$.

$$\mathbb{P}(M \le t) = \mathbb{P}(\max\{T_1, T_2, \dots, T_n\} \le t) = \mathbb{P}(\{T_1 \le t\} \cap \{T_2 \le t\} \cap \dots \cap \{T_n \le t\}) \overset{\text{independence}}{=} \mathbb{P}(T_1 \le t) \cdot \mathbb{P}(T_2 \le t) \cdot \dots$$

$$\cdot \mathbb{P}(T_n \le t) = F_{T_1}(t) \cdot F_{T_2}(t) \cdot \dots \cdot F_{T_n}(t) = (1 - e^{-t})^n \Rightarrow$$

$$\Rightarrow F_M(t) = (1 - e^{-t})^n.$$

We want to find the median of the distribution of M i.e. the number m that satisfies: $\mathbb{P}(M \le m) \ge \frac{1}{2}$ and $\mathbb{P}(M \ge m) \ge \frac{1}{2}$ and as $\mathbb{P}(M \le x) + \mathbb{P}(M \ge x) = 1$ $(\forall) x \in \mathbb{R}$, we then have $\mathbb{P}(M \le m) = \frac{1}{2}$.

That means $F_M(m) = \frac{1}{2} \Rightarrow (1 - e^{-m})^n = \frac{1}{2} \Rightarrow 1 - e^{-m} = \left(\frac{1}{2}\right)^{\frac{1}{n}} \Rightarrow e^{-m} = 1 - \left(\frac{1}{2}\right)^{\frac{1}{n}} \Rightarrow$

$$\Rightarrow -m = \log\left(1 - \left(\tfrac{1}{2}\right)^{\frac{1}{n}}\right) \Rightarrow \boxed{m = -\log\left(1 - \left(\tfrac{1}{2}\right)^{\frac{1}{n}}\right)}$$

As n becomes large, we can use the estimation $\alpha^{\frac{1}{n}} - 1 = e^{\frac{1}{n}\log\alpha} - 1 \approx \frac{1}{n}\log\alpha$, so m becomes:

$$m = -\log\left(1 - e^{\frac{1}{n}\log\frac{1}{2}}\right) \approx -\log\left(-\frac{1}{n}\log\frac{1}{2}\right) = -\log\left(\frac{1}{n}\log 2\right) = -\log\frac{1}{n} - \log(\log 2))$$

$m \approx \log n - \log(\log 2))$, which we can write as $m \approx \log n + c$, where c is a constant equal to $-\log(\log 2))$. Therefore, m grows as fast as $\log n$ does as n becomes large (we can ignore the constant c).

3. Let U and V be independent r.vs. with $U, V \sim U[0,1]$.

Then, $f_U(x) = f_V(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$ and

$$F_U(x) = F_V(x) = \begin{cases} x, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

The quadratic equation $x^2 + 2Ux + V = 0$ has two real solutions if $\Delta = 4U^2 - 4V \ge 0 \Rightarrow$

$\Rightarrow U^2 - V \ge 0 \Rightarrow U^2 \ge V \Rightarrow -U \le V \le U.$

We have:

$$\mathbb{P}(\text{two real solutions}) = \mathbb{P}(-U \le V \le U) = \mathbb{P}(V \le U) - \mathbb{P}(V \le -U) \overset{\text{independence}}{=} F_V(U) - F_V(-U).$$

$F_V(U) = \begin{cases} U, & \text{if } 0 \le U \le 1 \\ 0, & \text{otherwise} \end{cases} = U$ (as $U \sim U[0,1]$, so the probability that $0 \le U \le 1$ is 1)

$F_V(-U) = \begin{cases} -U, & \text{if } 0 \le -U \le 1 \Leftrightarrow -1 \le U \le 0 \text{ (false)} \\ 0, & \text{otherwise} \end{cases} = 0$

Then, $\boxed{\mathbb{P}(\text{two real solutions}) = U}$ (if $U = k$, with $k \in [0,1] \Rightarrow \mathbb{P}(\text{two real solutions}) = k$)

4. A fair die is thrown n times.

Let X denote the number of times we obtain a 6 from the throw. Therefore, as each throw X_i, $i \in \{1,2,\dots,n\}$ has a Bernoulli distribution with parameter $\frac{1}{6}$ ($X_i \sim Ber(\frac{1}{6})$), then, as $X = X_1 + X_2 + \dots + X_n$, $X \sim Bin(n, \frac{1}{6})$.

Then $\mathbb{E}(X) = \frac{n}{6}$ and $Var(X) = n \cdot \frac{1}{6} \cdot (1 - \frac{1}{6}) \Rightarrow Var(X) = \frac{5n}{36}$.

From Chebyshev's inequality, we have:

$$\mathbb{P}\left(|X - \mathbb{E}(X)| > \sqrt{n} \right) \leq \frac{Var(X)}{n} \Rightarrow \mathbb{P}\left(|X - \tfrac{n}{6}| > \sqrt{n} \right) \leq \frac{5}{36} \Rightarrow$$

$$\Rightarrow \mathbb{P}\left(|X - \tfrac{n}{6}| \leq \sqrt{n} \right) = 1 - \mathbb{P}\left(|X - \tfrac{n}{6}| > \sqrt{n} \right) \geq 1 - \frac{5}{36} = \frac{31}{36} \Rightarrow$$

$$\Rightarrow \mathbb{P}\left(-\sqrt{n} \leq X - \tfrac{n}{6} \leq \sqrt{n} \right) = \frac{31}{36} \Rightarrow \underline{\mathbb{P}\left(\tfrac{n}{6} - \sqrt{n} \leq X \leq \tfrac{n}{6} + \sqrt{n} \right) \geq \frac{31}{36}}.$$

5. Let $X_1, X_2, \dots, X_n$ be i.i.d. r.vs. from a distribution with mean μ and variance σ^2. Then, we have the sample mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ with $\mathbb{E}(\bar{X}_n) = \mu$ and $Var(\bar{X}_n) = \frac{\sigma^2}{n}$.

We want to find n such that

$$\mathbb{P}\left(|\bar{X}_n - \mu| < 2\sigma \right) \geq 0.99$$

the probability in a specific point is 0

Chebyshev's Inequality

$$\mathbb{P}\left(|\bar{X}_n - \mu| < 2\sigma \right) = 1 - \mathbb{P}\left(|\bar{X}_n - \mu| > 2\sigma \right) = 1 - \mathbb{P}\left(|\bar{X}_n - \mathbb{E}(\bar{X}_n)| > 2\sigma \right) \geq 1 - \frac{Var(\bar{X}_n)}{4\sigma^2} =$$

$$= 1 - \frac{\sigma^2}{4n\sigma^2} = 1 - \frac{1}{4n}.$$

if we have $1 - \frac{1}{4n} \geq 0.99 \Leftrightarrow 1 - \frac{1}{4n} \geq \frac{99}{100} \Leftrightarrow 100n - 25 \geq 99n \Leftrightarrow n \geq 25$,

then $\mathbb{P}\left(|\bar{X}_n - \mu| < 2\sigma \right) \geq 1 - \frac{1}{4n} \geq 0.99$, so for our inequality to hold, we need that $\boxed{n \geq 25}$.

6. A fair coin is tossed $(n+1)$ times $\Rightarrow \mathbb{P}(Heads) = \mathbb{P}(Tails) = \frac{1}{2}$.

Let $A_i = \begin{cases} 1, & \text{if the } i^{th} \text{ and } (i+1)^{st} \text{ outcomes are both H} \\ 0, & \text{otherwise} \end{cases}$, for $1 \leq i \leq n$.

(a) $\mathbb{P}(A_i = 1) = \frac{1}{4}$ (HH)

$\mathbb{P}(A_i = 0) = \frac{3}{4}$ (HT, TH, TT)

$\mathbb{E}(A_i) = 0 \cdot \mathbb{P}(A_i = 0) + 1 \cdot \mathbb{P}(A_i = 1) \Rightarrow \boxed{\mathbb{E}(A_i) = \frac{1}{4}}$, for all $1 \leq i \leq n$

$\mathbb{E}(A_i^2) = 0^2 \cdot \mathbb{P}(A_i = 0) + 1^2 \cdot \mathbb{P}(A_i = 1) \Rightarrow \mathbb{E}(A_i^2) = \frac{1}{4}$

$\mathbb{E}^2(A_i) = \frac{1}{16}$

$Var(A_i) = \mathbb{E}(A_i^2) - \mathbb{E}^2(A_i)$

$\Rightarrow \boxed{Var(A_i) = \frac{3}{16}}$, for all $1 \leq i \leq n$.

(b) We want to find $\mathrm{cov}(A_i, A_j)$, for $i, j \in \{1,2,\dots,m\}$, $i \neq j$

Case 1: $|i-j| = 1 \Rightarrow A_i$ and A_j are not independent

$$\mathrm{cov}(A_i, A_j) = \mathbb{E}(A_i A_j) - \mathbb{E}(A_i)\,\mathbb{E}(A_j)$$

$$\mathbb{E}(A_i A_j) = 0 \cdot 0 \cdot \mathbb{P}(A_i = 0, A_j = 0) + 0 \cdot 1 \cdot \mathbb{P}(A_i = 0, A_j = 1) + 1 \cdot 0 \cdot \mathbb{P}(A_i = 1, A_j = 0) + 1 \cdot 1 \cdot \mathbb{P}(A_i = 1, A_j = 1)$$

$\mathbb{E}(A_i A_j) = \mathbb{P}(A_i = 1, A_j = 1)$ (HHH is the only configuration that satisfies this)

$$\left.\begin{array}{l} \mathbb{E}(A_i A_j) = \frac{1}{8} \\ \mathbb{E}(A_i) = \mathbb{E}(A_j) = \frac{1}{4} \end{array}\right| \Rightarrow \boxed{\mathrm{cov}(A_i, A_j) = \frac{1}{16}}$$

Case 2: $|i-j| > 1 \Rightarrow A_i$ and A_j are independent $\Rightarrow \boxed{\mathrm{cov}(A_i, A_j) = 0}$

(c) Let $M = A_1 + A_2 + \dots + A_m$ denote the number of occurences of the motif HH in the sequence.

$$\mathbb{E}(M) = \mathbb{E}(A_1 + A_2 + \dots + A_m) = \mathbb{E}(A_1) + \mathbb{E}(A_2) + \dots + \mathbb{E}(A_m) \Rightarrow \boxed{\mathbb{E}(M) = \frac{m}{4}}$$

Now, we'll prove by induction that $\mathrm{var}(M) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{cov}(A_i, A_j)$, for all $n \geq 1$.

Base case: $P(1)$: $\mathrm{var}(A_1) = \mathrm{cov}(A_1, A_1) = \mathbb{E}(A_1^2) - \mathbb{E}^2(A_1)$, true

Inductive step

IH: We know that $P(n)$: $\mathrm{var}(A_1 + A_2 + \dots + A_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{cov}(A_i, A_j)$ and we want to prove

$$P(n+1): \ \mathrm{var}(A_1 + A_2 + \dots + A_{n+1}) = \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \mathrm{cov}(A_i, A_j)$$

$$\mathrm{var}(A_1 + A_2 + \dots + A_n + A_{n+1}) \overset{\text{property of var}}{=} \mathrm{var}(A_1 + A_2 + \dots + A_n) + \mathrm{var}(A_{n+1}) + 2\,\mathrm{cov}(A_1 + A_2 + \dots + A_n, A_{n+1}) = (\text{IH})$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{cov}(A_i, A_j) + \mathrm{cov}(A_{n+1}, A_{n+1}) + 2X, \text{ where } X = \mathrm{cov}(A_1 + A_2 + \dots + A_n, A_{n+1})$$

$$X = \mathbb{E}(A_1 A_{n+1} + A_2 A_{n+1} + \dots + A_n A_{n+1}) - \left(\mathbb{E}(A_1) + \mathbb{E}(A_2) + \dots + \mathbb{E}(A_n)\right)\mathbb{E}(A_{n+1}) =$$

$$= \mathrm{cov}(A_1, A_{n+1}) + \mathrm{cov}(A_2, A_{n+1}) + \dots + \mathrm{cov}(A_n, A_{n+1})$$

Therefore,

$$\mathrm{var}(A_1 + A_2 + \dots + A_{n+1}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{cov}(A_i, A_j) + \mathrm{cov}(A_{n+1}, A_{n+1}) + 2\left(\sum_{i=1}^{n} \mathrm{cov}(A_i, A_{n+1})\right) =$$

$$= \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \mathrm{cov}(A_i, A_j) \Rightarrow \text{the inductive step is proven}$$

Therefore, we proved that

$$\mathrm{var}(M) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{cov}(A_i, A_j) = \sum_{i=1}^{n} \mathrm{var}(A_i) + 2(n-1) \cdot \frac{1}{16} = \frac{3m}{16} + \frac{m-1}{8} \Rightarrow \boxed{\mathrm{var}(M) = \frac{5m-2}{16}}$$

(d) Let $B_i = \begin{cases} 1, & \text{if the } i^{th} \text{ is T and the } (i+1)^{st} \text{ is H} \\ 0, & \text{otherwise} \end{cases}$

We observe that $\mathbb{E}(B_i) = \frac{1}{4}$, $\text{var}(B_i) = \frac{3}{16}$ (similar to A_i)

$\mathbb{E}(B_i B_j) = 0$ if $|i-j| = 1$ (as we cannot two TH in a sequence of 3 consecutive outcomes) $\Rightarrow$

$\Rightarrow \text{cov}(B_i, B_j) = -\mathbb{E}(B_i)\mathbb{E}(B_j) = -\frac{1}{16}$, if $|i-j| = 1$

$\text{cov}(B_i, B_j) = 0$ if $|i-j| > 1$, as B_i and B_j are independent.

Let $N = B_1 + B_2 + \dots + B_n$ denote the number of occurences of the motif TH in the sequence.

Then, $\mathbb{E}(N) = \mathbb{E}(B_1 + B_2 + \dots + B_n) = \mathbb{E}(B_1) + \mathbb{E}(B_2) + \dots + \mathbb{E}(B_n) \Rightarrow \boxed{\mathbb{E}(N) = \frac{n}{4}}$

$$\text{var}(N) = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{cov}(B_i, B_j) = \sum_{i=1}^{n} \text{var}(B_i) - 2 \cdot (n-1) \cdot \frac{1}{16} = \frac{3n}{16} - \frac{2n-2}{16} \Rightarrow \boxed{\text{Var}(N) = \frac{n+2}{16}}$$

7. Let $a, b, p \in (0,1)$.

Let $X_1, X_2, \dots, X_n$ i.i.d. r.vs. with $X_i \sim Br(p)$, $i \in \{1, 2, \dots, n\}$.

Let $Y = X_1 + X_2 + \dots + X_n$. Then, $Y \sim Bin(n, p)$, so the probability distribution function of Y is $\mathbb{P}(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, 2, \dots, n$

$\mathbb{E}(X_i) = p$, for all $i \in \{1, 2, \dots, n\}$

From the weak law of large numbers, we have that

$(\forall)\ \varepsilon > 0$, $(*)\ \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > \varepsilon\right) \to 0$ as $n \to \infty$ and

$(**)\ \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| \le \varepsilon\right) \to 1$ as $n \to \infty$.

$$L = \lim_{n\to\infty} \sum_{r \in \mathbb{N}} \binom{n}{r} p^r (1-p)^{n-r} = \lim_{n\to\infty} \sum_{\substack{r \in \mathbb{N} \\ an < r < bn}} \mathbb{P}(Y = r) = \lim_{n\to\infty} \mathbb{P}(an < Y < bn) =$$

$$= \lim_{n\to\infty} \mathbb{P}\left(a \le \frac{Y}{n} \le b\right) = \lim_{n\to\infty} \mathbb{P}\left(a - p \le \frac{Y}{n} - p \le b - p\right)$$

(i) $p < a \Rightarrow a - p > 0$. Let $\varepsilon = a - p > 0$, then

$$L = \lim_{n\to\infty} \mathbb{P}\left(\varepsilon \le \frac{Y}{n} - p \le b - a + \varepsilon\right) \le \lim_{n\to\infty} \mathbb{P}\left(\varepsilon \le \frac{Y}{n} - p\right) \le \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| \ge \varepsilon\right) \overset{(*)}{=} 0$$

$\Rightarrow \boxed{L = 0}$ (as $L \in [0,1]$)

(ii) $a < p < b \Rightarrow p - a > 0, b - p > 0$. Let $\varepsilon = \min\{p-a, b-p\} > 0$, then

$$L = \lim_{n\to\infty} \mathbb{P}\left(a - p \le \frac{Y}{n} - p \le b - p\right) \ge \lim_{n\to\infty} \mathbb{P}\left(-\varepsilon \le \frac{Y}{n} - p \le \varepsilon\right) = \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| \le \varepsilon\right) \overset{(**)}{=}$$

$\Rightarrow \boxed{L=1}$ (as $L \in [0,1]$)

(iii) $b < p \Rightarrow p - b > 0$. Let $\varepsilon = p - b > 0$, then

$$L = \lim_{n\to\infty} \mathbb{P}\left(a - b - \varepsilon \leq \frac{Y}{n} - p \leq -\varepsilon\right) \leq \lim_{n\to\infty} \mathbb{P}\left(\frac{Y}{n} - p \leq -\varepsilon\right) \leq \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| \geq \varepsilon\right) \overset{(*)}{=} 0$$

$\Rightarrow \boxed{L=0}$ (as $L \in [0,1]$)

We switched from strict signs $(>, <)$ to $(\geq, \leq)$ and vice-versa throughout the proof as we work with continuous probabilities and functions, so in every point the probability is 0.

Therefore,

$$L = \lim_{n\to\infty} \sum_{\substack{r \in \mathbb{N}: \\ an < r < bn}} \binom{n}{r} p^r (1-p)^{n-r} = \begin{cases} 1, & \text{if } a < p < b \\ 0, & \text{if } p < a \text{ or } b < p. \end{cases}$$