# AI701 Assignment 1

Sultan Rizky Hikmawan Madjid (20214196)

Due: September 21, 2021

1. For a random variable $X$ denoting the number of heads among $n$ coin tosses which follows the binomial distribution with parameter $\theta$,

$$\mathbb{P}(X = x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

   (a) Given that $\theta$ itself is a random variable following the beta distribution with Probability Density Function (PDF) with parameters $a, b > 0$,

$$f(\theta; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{0\le\theta\le1}$$

   for gamma function $\Gamma(t) = \int_0^\infty z^{t-1}e^{-z}\,\mathrm{d}z$, the marginal likelihood of $x$ under this setting can be expressed as

$$\begin{aligned}
\mathbb{P}(X = x; a, b) &= \int_0^1 \mathbb{P}(X = x|\theta)f(\theta; a, b)\,\mathrm{d}\theta \\
&= \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \theta^{x+a-1}(1-\theta)^{n+b-x-1}\,\mathrm{d}\theta \\
&= \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left(\cancelto{0}{\left[\frac{\theta^{x+a}(1-\theta)^{n+b-x-1}}{x+a}\right]_0^1} + \frac{n+b-x-1}{x+a}\int_0^1 \theta^{x+a}(1-\theta)^{n+b-x-2}\,\mathrm{d}\theta\right) \\
&= \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{(n+b-x-1)\cdot(n+b-x-2)\cdot\ldots\cdot1}{(a+b+n-2)\cdot(a+b+n-3)\cdot\ldots\cdot(x+a)}\int_0^1 \theta^{a+n+b-2}\,\mathrm{d}\theta \\
&= \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{(n+b-x-1)\cdot(n+b-x-2)\cdot\ldots\cdot1}{(a+b+n-1)\cdot(a+b+n-2)\cdot\ldots\cdot(x+a)} \\
&= \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(n+b-x)\Gamma(x+a)}{\Gamma(a+b+n)}
\end{aligned}$$

   (b) Assuming that we toss the coin $n = 10$ times and see $x = 9$ heads, for two prior distributions $\mathcal{M}_1$ and $\mathcal{M}_2$ for $\theta$ defined as:

$$\mathcal{M}_1 : \mathbb{P}(\theta) = \delta_{1/2}(\theta), \quad \mathcal{M}_2 : f(\theta; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{\{0\le\theta\le1\}}.$$

   Fixing $a = b = 1$, we'll have that Bayes factor

$$\begin{aligned}
\frac{\mathbb{P}(X = x|\mathcal{M}_1)}{\mathbb{P}(X = x|\mathcal{M}_2)} &= \frac{\int \mathbb{P}(X = x|\theta)\mathbb{P}(\theta|\mathcal{M}_1)\,\mathrm{d}\theta}{\int \mathbb{P}(X = x|\theta)\mathbb{P}(\theta|\mathcal{M}_2)\,\mathrm{d}\theta} \\
&= \frac{\binom{n}{x}(1/2)^x(1-1/2)^{n-x}}{\binom{n}{x}\int_0^1 \theta^x(1-\theta)^{n-x}\,\mathrm{d}\theta} \\
&= \frac{(1/2)^n\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} \\
&= \frac{(1/2)^{10}(10+1)!}{9!(10-9)!} = \frac{55}{512} = 0.107421875
\end{aligned}$$

2. For a set $E$ and its collection of subsets $\mathcal{A}$, defining $\sigma(\mathcal{A})$ as the smallest $\sigma$-algebra containing $\mathcal{A}$ or the intersection of all $\sigma$-algebras containing $\mathcal{A}$,

   (a) Given $E = \{1, 2, 3, 4\}$ and $\mathcal{A} = \{\{1\}, \{2\}\}$, from the definition of $\sigma$-algebra itself, given $\mathcal{A} = \{A, B\}$, we have that $\sigma(\mathcal{A}) = \{\varnothing, E, A, B, E \setminus A, E \setminus B, A \cup B, E \setminus (A \cup B)\}$. In other words,

$$\sigma(\mathcal{A}) = \{\varnothing, \{1\}, \{2\}, \{1, 2\}, \{3, 4\}, \{2, 3, 4\}, \{1, 3, 4\}, \{1, 2, 3, 4\}\}$$

   (b) For a set $E$ and its collection of subsets $\mathcal{A}, \mathcal{C}$,

      1. *Proof.* Given $\mathcal{A} \subset \mathcal{C}$, from the definition of $\sigma$-algebra generated by $\mathcal{C}$, $\sigma(\mathcal{C})$, which is the smallest $\sigma$-algebra on $E$ that contains $\mathcal{C}$, we have that $\mathcal{C} \subset \sigma(\mathcal{C})$. Intuitively, this implies that $\mathcal{A} \subset \mathcal{C} \subset \sigma(\mathcal{C})$, or in short, $\mathcal{A} \subset \sigma(\mathcal{C})$. In other words, $\sigma(\mathcal{C})$ is a $\sigma$-algebra which contains the sets in $\mathcal{A}$. Thus, from the definition of $\sigma$-algebra generated by $\mathcal{A}$, which is the smallest $\sigma$-algebra on $E$ that contains $\mathcal{A}$, we have that $\mathcal{A} \subset \sigma(\mathcal{A}) \subset \sigma(\mathcal{C})$, or in short $\sigma(\mathcal{A}) \subset \sigma(\mathcal{C})$. $\square$

      2. *Proof.* Given $\mathcal{A} \subset \sigma(\mathcal{C})$ and $\mathcal{C} \subset \sigma(\mathcal{A})$, it can be noticed that given $\mathcal{A} \subset \sigma(\mathcal{C})$, from the definition of $\sigma$-algebra generated by $\mathcal{A}$, which is the smallest $\sigma$-algebra on $E$ that contains $\mathcal{A}$, we have that $\mathcal{A} \subset \sigma(\mathcal{A}) \subset \sigma(\mathcal{C})$, or in short, $\sigma(\mathcal{A}) \subset \sigma(\mathcal{C})$. Similarly, given $\mathcal{C} \subset \sigma(\mathcal{A})$, from the definition of $\sigma$-algebra generated by $\mathcal{C}$, which is the smallest $\sigma$-algebra on $E$ that contains $\mathcal{C}$, we have that $\mathcal{C} \subset \sigma(\mathcal{C}) \subset \sigma(\mathcal{A})$, or in short, $\sigma(\mathcal{C}) \subset \sigma(\mathcal{A})$. Since we have $\sigma(\mathcal{C}) \subset \sigma(\mathcal{A})$ and $\sigma(\mathcal{A}) \subset \sigma(\mathcal{C})$, therefore, it implies that $\sigma(\mathcal{A}) = \sigma(\mathcal{C})$. $\square$

      3. *Proof.* Given $\mathcal{A} \subset \mathcal{C} \subset \sigma(\mathcal{A})$, from 2(b)1, we have that given $\mathcal{A} \subset \mathcal{C}$, $\sigma(\mathcal{A}) \subset \sigma(\mathcal{C})$, which further implies that $\mathcal{A} \subset \sigma(\mathcal{A}) \subset \sigma(\mathcal{C})$. Now, from 2(b)2, since we have both $\mathcal{A} \subset \sigma(\mathcal{C})$ and $\mathcal{C} \subset \sigma(\mathcal{A})$, it means that $\sigma(\mathcal{A}) = \sigma(\mathcal{C})$, completing our proof. $\square$

3. For Moment Generating Function (MGF) of a $\mathbb{R}$-valued random variable $X$ defined as

$$M_X(t) = \mathbb{E}\left[e^{tX}\right]$$

(a) For a standard normal random variable $X$ with PDF $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, we have that it's Moment Generating Function (MGF) is given by:

$$\begin{aligned} M_X(t) = \mathbb{E}\left[e^{tX}\right] &= \int_{-\infty}^{\infty} e^{tx} f_X(x)\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx-\frac{x^2}{2}}\,\mathrm{d}x \\ &= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2}\,\mathrm{d}x \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

(b) Defining $X_1, ..., X_n$ for $n \in \mathbb{N}$ as i.i.d. $\mathbb{R}$-valued random variables with $M_{X_1}(t)$ defined as the MGF of $X_1$, defining $Y_n := c(X_1 + X_2 + ... + X_n)$, it's MGF can be mathematically expressed as:

$$\begin{aligned} M_{Y_n}(t) &= \mathbb{E}\left[e^{tY_n}\right] = \mathbb{E}\left[e^{tc(X_1+X_2+...+X_n)}\right] \\ &= \prod_{i=1}^{n} \mathbb{E}\left[e^{tcX_i}\right] = \left(\mathbb{E}\left[e^{tcX_1}\right]\right)^n = \left(\mathbb{E}\left[e^{ctX_1}\right]\right)^n \\ &= (M_{X_1}(ct))^n \end{aligned}$$

(c) *Proof.* For i.i.d. $\mathbb{R}$-valued random variables $X_i$ with mean $\mu$ and variance $\sigma^2$, we first define for $n \in \mathbb{N}$ that

$$Y_i := \frac{X_i - \mu}{\sigma}, \quad Z_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i.$$

Now, to simplify our calculations, we are going to use random variables $Y_i$ which satisfy $\mathbb{E}[Y_i] = \mathbb{E}\left[\frac{X_i-\mu}{\sigma}\right] = 0$ and $\mathbb{E}\left[Y_i^2\right] = \mathbb{E}\left[\frac{(X_i-\mu)^2}{\sigma^2}\right] = \frac{\mathrm{Var}[X_i]}{\sigma^2} = 1$. The MGF of $Z_n$ can be calculated as:

$$\begin{aligned} M_{Z_n}(t) &= \mathbb{E}\left[e^{tZ_n}\right] \\ &= \mathbb{E}\left[\exp\left(\frac{t}{\sqrt{n}}\left(\sum_{i=1}^{n} Y_i\right)\right)\right] \\ &= \prod_{i=1}^{n} \mathbb{E}\left[e^{tY_i/\sqrt{n}}\right] = \left(M_{Y_i}\left(\frac{t}{\sqrt{n}}\right)\right)^n \end{aligned}$$

From Taylor Series expansion, by denoting $o(t^2)$ as a function that satisfies $\lim_{t\to 0} o(t^2)/t^2 = 0$, it is possible to express $M_{Z_n}(t)$ as

$$\begin{aligned} M_{Z_n}(t) &= \left(M_{Y_i}\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(M_{Y_i}(0) + \frac{\mathbb{E}[Y_i]\,t}{\sqrt{n}} + \frac{\mathbb{E}\left[Y_i^2\right]t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \\ &= \left(1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \end{aligned}$$

Now, taking the limit as $n \to \infty$ and neglecting the higher-order terms represented by $o(\frac{t^2}{n})$ since it converges to 0 faster than the other existing terms, it is possible to deduce that

$$\begin{aligned} \lim_{n\to\infty} M_{Z_n}(t) &= \lim_{n\to\infty} \left(1 + \frac{t^2}{2n}\right)^n \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

which is the MGF of a standard normal distribution as derived in 3a. Thus, $Z_n$ converges in distribution to a standard normal distribution as $n \to \infty$. □

4. For a random variable $X$ with finite mean $\mu$ and variance $\sigma^2$ and its i.i.d. copies $X_1, ..., X_n$, defining estimator for $\sigma$:

$$\hat{\sigma}_n^2 := \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2, \quad \bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(a) *Proof.* From the Weak Law of Large Numbers, it can be noticed that as $n \to \infty$, we have that $\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mathbb{E}[X]$ and similarly, $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p} \mathbb{E}\left[X^2\right]$. From the fact that

$$\hat{\sigma}_n^2 := \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2$$

Using both propositions deduced from the Weak Law of Large Numbers and the equation above, it can be implied that as $n \to \infty$, $\hat{\sigma}_n^2 \xrightarrow{p} \mathbb{E}\left[X^2\right] - \left(\mathbb{E}[X]\right)^2 = \mathrm{Var}[X] = \sigma^2$ or in short, $\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2$. Thus, $\hat{\sigma}_n^2$ is consistent. □

(b) *Proof.* Using the expanded form of $\hat{\sigma}_n^2$ derived in 4a, it is possible to express that

$$\begin{aligned}
\mathbb{E}\left[\hat{\sigma}_n^2\right] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[\bar{X}_n^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{Var}[X_i] + \left(\mathbb{E}[X_i]\right)^2\right) - \left(\mathrm{Var}\left[\bar{X}_n\right] + \left(\mathbb{E}\left[\bar{X}_n\right]\right)^2\right) \\
&= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 \\
&= \frac{n-1}{n}\sigma^2 \neq \sigma^2
\end{aligned}$$

Due to the fact that $\mathbb{E}\left[\hat{\sigma}_n^2\right] \neq \sigma^2$, $\hat{\sigma}_n^2$ is biased. □

(c) In order to make $\hat{\sigma}_n^2$ unbiased, it is possible to redefine the estimator of $\sigma$ as follows:

$$\hat{\sigma}_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2, \quad \bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

By this way, we are going to have an estimator $\hat{\sigma}_n^2$ which is *consistent* from the fact that given $n \to \infty$, $\hat{\sigma}_n^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2\right) \xrightarrow{p} (1) \cdot \sigma^2$, or in short, $\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2$ and *unbiased*, using the fact that

$$\begin{aligned}
\mathbb{E}\left[\hat{\sigma}_n^2\right] &= \frac{n}{n-1}\left(\sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2\right) \\
&= \sigma^2
\end{aligned}$$

5. *Proof.* For a sequence of $\mathbb{R}$-valued random variables $(X_n)_{n\geq 1}$ with distribution functions $(F_n)_{n\geq 1}$ and a $\mathbb{R}$-valued random variable $X$ with distribution $F$, given $X_n \xrightarrow{d} c$, where $c$ is a constant, from the definition of Convergence in Distribution that is being normally presented by the notation $X_n \xrightarrow{d} X$, where for all $x \in \mathbb{R}$, $\lim_{n\to\infty} F_n(x) = F(x)$, given $X = c$, we'll have that $F(x) = 0$ if $x < c$ and $F(x) = 1$ if $x \geq c$. Here, $F(x)$ is continuous everywhere except at $c$. This implies that for any $\epsilon > 0$, we have that

$$\lim_{n\to\infty} F_n(c-\epsilon) = \lim_{n\to\infty} [\mathbb{P}(X_n \leq c-\epsilon)] = 0,$$
$$\lim_{n\to\infty} F_n\Big(c+\frac{\epsilon}{2}\Big) = \lim_{n\to\infty} \Big[\mathbb{P}\Big(X_n \leq c+\frac{\epsilon}{2}\Big)\Big] = 1.$$

Naturally, we have that $\lim_{n\to\infty} \mathbb{P}(|X_n - c| > \epsilon) \geq 0$ for every $\epsilon > 0$. However, it can be noticed that

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(|X_n - c| > \epsilon) &= \lim_{n\to\infty} [\mathbb{P}(X_n < c-\epsilon) + \mathbb{P}(X_n > c+\epsilon)] \\
&= \lim_{n\to\infty} [\mathbb{P}(X_n < c-\epsilon)] + \lim_{n\to\infty} [\mathbb{P}(X_n > c+\epsilon)] \\
&= 0 + \lim_{n\to\infty} [\mathbb{P}(X_n > c+\epsilon)] \\
&\leq \lim_{n\to\infty} \Big[\mathbb{P}\Big(X_n > c+\frac{\epsilon}{2}\Big)\Big] \\
&= 1 - \lim_{n\to\infty} \Big[F_n\Big(c+\frac{\epsilon}{2}\Big)\Big] \\
&= 0
\end{aligned}$$

Therefore, from both inequalities, we have that $\lim_{n\to\infty} \mathbb{P}(|X_n - c| > \epsilon) = 0$ for every $\epsilon > 0$, or equivalently, $X_n \xrightarrow{p} c$. □