Probability and Statistics
Supervision 1

Samuel Lee

Question 1

A σ-algebra is on some set Ω is a collection $\mathcal{F}$ of subsets of X such that

1. $A \in \mathcal{F} \implies \Omega \setminus A \in \mathcal{F}$
2. $\Omega \in \mathcal{F}$
3. $A_1, A_2, \dots \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

If $\mathcal{B}$ is some collection of subsets of Ω, then the σ-algebra **generated by** $\mathcal{B}$, denoted by $\sigma(\mathcal{B})$, is the smallest σ-algebra on Ω which contains $\mathcal{B}$.

In this case, $\Omega = S = \{a, b, c, d, e\}$ and $\mathcal{B} = \{\{a, b, c\}\{c, d, e\}\}$. The σ-algebra $\mathcal{A}$ generated by $\mathcal{B}$ is $\{\varnothing, \{a, b, c\}, \{c, d, e\}, \{d, e\}, \{a, b\}, \{a, b, d, e\}, \{c\}, \mathcal{S}\}$.

Question 2

(a)

Let I represent the event where a person is infected and let P represent the event where a person tests positive. We know that $\Pr(I) = 0.01$, $\Pr(P|I) = 0.9$, and $\Pr(P|I^\complement) = 0.05$. We want to find $\Pr(I|P)$.

$$\begin{aligned}
\Pr(P|I) &= \frac{\Pr(P \cap I)}{\Pr(I)} = \frac{\Pr(P \cap I)}{0.01} = 0.9 \\
\Pr(P \cap I) &= 0.009 \\
\Pr(P) &= \Pr(P|I) \times \Pr(I) + \Pr(P|I^\complement) \times \Pr(I^\complement) \\
&= 0.9 \times 0.01 + 0.1 \times 0.99 = 0.108 \\
\Pr(I|P) &= \frac{\Pr(P \cap I)}{\Pr(P)} = \frac{0.009}{0.108} = \frac{1}{12}
\end{aligned}$$

(b)

Now, we have that $\Pr(T|I) = 0.3$ and $\Pr(T|I^\complement) = 0.15$, where T is the event where a person takes the test. The probabilities $\Pr(P|I)$ and $\Pr(P|I^\complement)$ should now be $\Pr(P|I \cap T) = 0.9$ and

$\Pr(P|I^{\complement} \cap T) = 0.05$. We want to find $\Pr(I|P \cap T)$.

$$\begin{aligned}
\Pr(T \cap I) &= \Pr(T|I) \times \Pr(I) \\
&= 0.3 \times 0.01 = 0.003 \\
\Pr(P \cap I \cap T) &= \Pr(P|I \cap T) \times \Pr(I \cap T) \\
&= 0.9 \times 0.003 = 0.0027 \\
\Pr(P \cap T) &= \Pr(I) \times \Pr(T|I) \times \Pr(P|T \cap I) + \Pr(I^{\complement}) \times \Pr(T|I^{\complement}) \times \Pr(P|T \cap I^{\complement}) \\
&= 0.01 \times 0.3 \times 0.9 + 0.99 \times 0.15 \times 0.005 = 0.007425 \\
\Pr(I|P \cap T) &= \frac{\Pr(I \cap P \cap T)}{\Pr(P \cap T)} \\
&= \frac{0.0027}{0.007425} = \frac{4}{11}
\end{aligned}$$

Intuitively, it is now more likely that one is infected given that the test is positive $\left(\frac{4}{11} > \frac{1}{12}\right)$ because there is self-selection into the test.

Question 3

X and Y are independent if their joint probability density function is the product of their marginal probability density functions. To get the marginal probability density functions,

$$\begin{aligned}
f_X(x) &= \int_0^1 f_{XY}(x,y)\mathrm{d}y \\
&= \int_0^1 c + \frac{x+y}{2}\mathrm{d}y \\
&= \left[\frac{2c+x}{2}y + \frac{y^2}{4}\right]_0^1 = \frac{4c+2x+1}{4}
\end{aligned}$$

with $f_Y(y)$ being the same as above but with y instead of x since the two are symmetric in this case. So we have

$$\begin{aligned}
f_X(x) \cdot f_Y(y) &= \frac{4c+2x+1}{4} \cdot \frac{4c+2y+1}{4} \\
&= \frac{16c^2 + 8cy + 4c + 8cx + 4xy + 2x + 4c + 2y + 1}{16}
\end{aligned}$$

and it is clear that the mess above is not equal to $f_{XY}(x,y)$. For one, there is no xy term in the joint density distribution. Thus X and Y are not independent.

The integral of a probability density function over its domain must be 1, so c must satisfy

$$\begin{aligned}
\int_0^1 f_X(x)\mathrm{d}x &= 1 \\
\int_0^1 \frac{4c+2x+1}{4}\mathrm{d}x &= 1 \\
\left[\frac{4c+1}{4}x + \frac{x^2}{4}\right]_0^1 &= 1 \\
\frac{4c+2}{4} &= 1 \\
c &= \frac{1}{2}
\end{aligned}$$

The expectation of X is thus

$$\begin{aligned}
\mathrm{E}(X) &= \int_0^1 x \cdot f_X(x)\mathrm{d}x \\
&= \int_0^1 \frac{2x^2+3x}{4}\mathrm{d}x \\
&= \left[\frac{x^3}{6} + \frac{3x^2}{8}\right]_0^1 = \frac{13}{24}
\end{aligned}$$

The conditional expectation of Y given that $X = \frac{1}{2}$ is

$$\begin{aligned}
\mathrm{E}\left(Y|X = \frac{1}{2}\right) &= \int_0^1 y \cdot f\left(\frac{1}{2}, y\right)\mathrm{d}y \\
&= \int_0^1 \frac{3y+2y^2}{4}\mathrm{d}y \\
&= \left[\frac{3y^2}{8} + \frac{y^3}{6}\right]_0^1 = \frac{13}{24}
\end{aligned}$$

Question 4

We have that the cumulative distribution function $F_Y(y)$ of Y satisfies

$$\begin{aligned}
F_Y(y) &= \Pr(Y \leq y) \\
&= \Pr(\cos(X) \leq y)
\end{aligned}$$

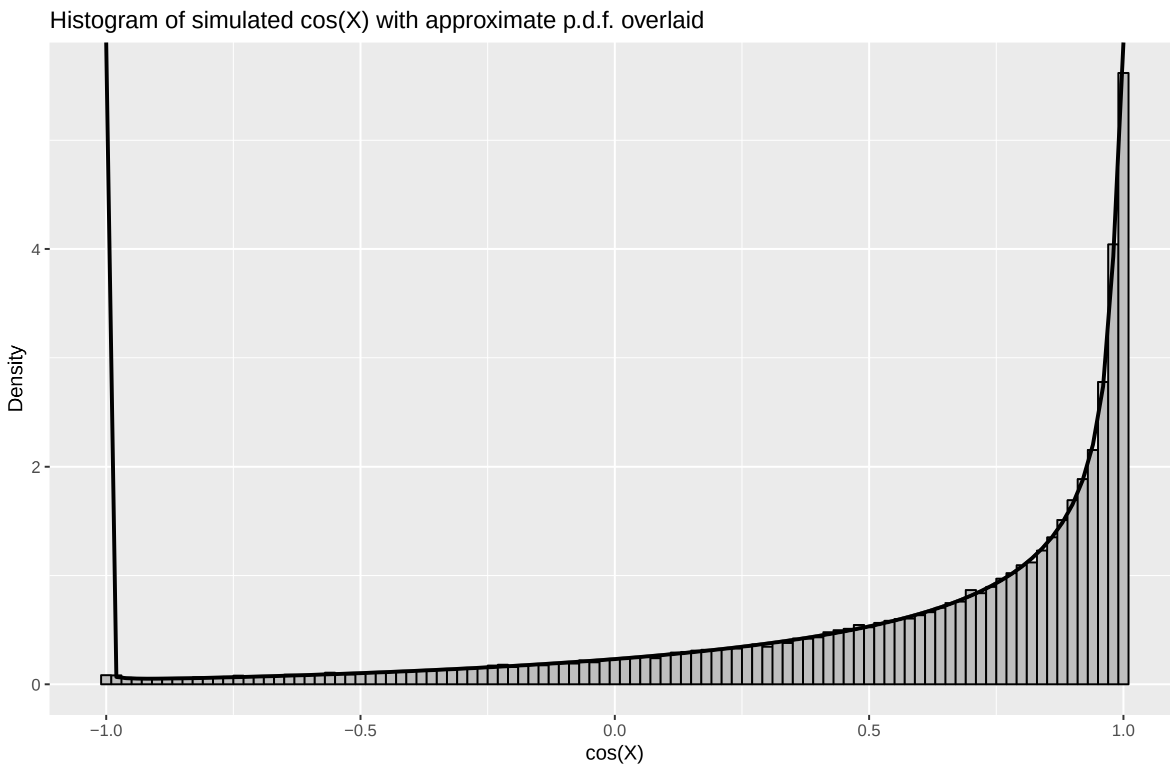
If $\cos(X) = c$, then $\cos(2\pi - X) = c$. Furthermore, we have that $\cos(X)$ is decreasing over $[k\pi, (k+1)\pi]$ for even k and increasing otherwise. Lastly, we have that $0 \leq \cos^{-1}(X) \leq \pi$. Therefore the above expression is

$$\begin{aligned}
\Pr(\cos(X) \leq y) &= \sum_{k=-\infty}^{\infty} \Pr\left[\underbrace{2k\pi + \cos^{-1}(y)}_{\text{decreasing region of } \cos(\cdot)} \leq X \leq \underbrace{(2k+2)\pi - \cos^{-1}(y)}_{\text{increasing region of } \cos(\cdot)}\right] \\
&= \sum_{k=-\infty}^{\infty} \left\{\Phi\left[2(k+1)\pi - \cos^{-1}(y)\right] - \Phi\left[2k\pi + \cos^{-1}(y)\right]\right\}
\end{aligned}$$

Taking the derivative with respect to y, we have

$$\begin{aligned}
f_Y(y) &= \sum_{k=-\infty}^{\infty} \left\{ \frac{1}{\sqrt{1-y^2}} \varphi[2(k+1)\pi - \cos^{-1}(y)] + \frac{1}{\sqrt{1-y^2}} \varphi[2k\pi + \cos^{-1}(y)] \right\} \\
&= \sum_{k=-\infty}^{\infty} \left\{ \frac{1}{\sqrt{1-y^2}} \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{[2(k+1)\pi - \cos^{-1}(y)]^2}{2}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{[2k\pi + \cos^{-1}(y)]^2}{2}} \right] \right\} \\
&= \frac{1}{\sqrt{2\pi(1-y^2)}} \sum_{k=-\infty}^{\infty} \left[e^{-\frac{[2(k+1)\pi - \cos^{-1}(y)]^2}{2}} + e^{-\frac{[2k\pi + \cos^{-1}(y)]^2}{2}} \right]
\end{aligned}$$

And if someone is good enough to find an analytical form for this expression, they will have found the probability density function for $\cos(X)$. As a consolation, below is the histogram of 100,000 simulated values of X and $\cos(X)$, with an approximation of the p.d.f. above overlaid (here the summation only runs from -1 to 1):



Question 5

It is clear that m_3 lies between m_1 and m_2 inclusive since it is the arithmetic mean, so we just have to show that m_1 is weakly greater than m_2. First we assume this isn't so, and $m_1 < m_2$. Since m_1 is the greatest lower bound of $\{x : F_X(x) > 1/2\}$, $m_1 < m_2$ implies that there is some x between m_1 and m_2 inclusive that satisfies $F_X(x) > 1/2$, or else $\{x : F_X(x) > 1/2\}$ is the empty set and there cannot be an infimum. By an analogous argument, there is some x between m_1 and m_2 inclusive that satisfies $F_X(x) < 1/2$.

If m_2 is the least upper bound of $\{x : F_X(x) < 1/2\}$ but not in the set, then $F_X(m_2) \geq 1/2$. But we have shown that there is some x between m_1 and m_2 where $F_X(x) > 1/2$. This means there is some point where F_X strictly decreases, which means it cannot be a valid cumulative distribution function[1]. Likewise, if m_2 is the least upper bound of $\{x : F_X(x) < 1/2\}$ which is also in the set, then $F_X(m_2) < 1/2$. But again, we found that there should be some x between m_1 and m_2 where $F_X(x) > 1/2$. And this is a contradiction for the same reasons as before. Therefore we cannot have $m_1 < m_2$, and it is at least not disproven that $m_1 \geq m_2$. We just need to find examples where $m_1 = m_2$ and $m_1 > m_2$ to show that $m_1 \geq m_2$ holds. The former is trivial, so we deal with the latter. If we have $F_X : \mathbb{N} \to \mathbb{R}$, where

$$F_X(x) = \begin{cases} \frac{1}{3} & \text{if } x = 1 \\ \frac{2}{3} & \text{if } x = 2 \\ 1 & \text{otherwise} \end{cases}$$

Then $\{x : F_X(x) > 1/2\} = \{2, 3, 4, \ldots\}$ and the infimum m_1 is 2. On the other hand $\{x : F_X(x) < 1/2\} = \{1\}$ and the supremum m_2 is 1.

$m_1 = m_2 = m_3$ when F_X is continuous and strictly monotonic. m_1 and m_2 will be equal to m where $F_X(m) = 1/2$. Suppose this is not so, and $F_X(m_1) > 1/2$ (if $F_X(m_1) < 1/2$ then m_1 is not a greatest lower bound; m is greater than m_1 and is still a lower bound). We know m exists because of continuity due to the intermediate value theorem. We just need to take $m_1' = \frac{m_1+m}{2}$ which will violate the definition of m_1 as the greatest lower bound: m_1' is smaller than m_1 (because by monotonicity of F_X the mean of m_1 and m must be less than m_1) but still in the set $\{x : F_X(x) > 1/2\}$ since $m_1' > m$. Hence $m_1 = m$. An analogous argument applies for m_2, and thus continuity and monotonicity are sufficient for $m_1 = m_2 = m_3$. The arguments for necessity can be found through counterexamples: an F_X which is equal to $\frac{1}{2}$ over a range $[a, b]$ (and therefore non-increasing over some range) and increasing everywhere else will have $m_1 = b$ and $m_2 = a$. The case where F_X is non-continuous has already been discussed above. Therefore continuity and monotonicity are necessary and sufficient conditions for $m_1 = m_2 = m_3$.

Question 6

We want to minimize

$$Q(\theta_1, \theta_2) = \mathrm{E}\left[\{(X - \theta_1)^2 - \theta_2\}^2\right]$$

We take the partial derivatives of $Q(\theta_1, \theta_2)$ and set them to zero:

$$\begin{aligned} \frac{\partial Q}{\partial \theta_1} &= -4 \cdot \mathrm{E}\left[\{(X - \theta_1)^2 - \theta_2\}(X - \theta_1)\right] = 0 \\ \frac{\partial Q}{\partial \theta_2} &= -2 \cdot \mathrm{E}\left[(X - \theta_1)^2 - \theta_2\right] = 0 \end{aligned}$$

We get from the second FOC:

$$\begin{aligned} \mathrm{E}\left[(X - \theta_1)^2\right] &= \theta_2 \\ \mathrm{E}\left[X^2 - 2X\theta_1 + \theta_1^2\right] &= \theta_2 \\ 1 + \theta_1^2 &= \theta_2 \end{aligned}$$

[1] The question says to consider the general case where F_X is not necessarily monotonic, but I will take this to mean strict monotonocity. I might just be wrong but I haven't been able to find a proof in a case where F_X is allowed to be non-monotonic even in the weak sense. In fact it seems easy to come up with non-monotonic functions where $m_1 < m_2$: a quadratic function could suffice as long as the domain is restricted from the left (needed for m_1 to exist).

From the first FOC:

$$
\begin{aligned}
\mathrm{E}\left[\{(X-\theta_1)^2-\theta_2\}(X-\theta_1)\right]&=0\\
\mathrm{E}\left[\{X^2-2X\theta_1+\theta_1^2-\theta_2\}(X-\theta_1)\right]&=0\\
\mathrm{E}\left[X^3-\theta_1X^2-2\theta_1X^2+2\theta_1^2X+\theta_1^2X-\theta_1^3-\theta_2X+\theta_1\theta_2\right]&=0\\
\mathrm{E}\left[X^3-3\theta_1X^2+(3\theta_1^2-\theta_2)X+\theta_1(\theta_2-\theta_1^2)\right]&=0\\
-3\theta_1+\theta_1(\theta_2-\theta_1^2)&=0\\
\theta_1(\theta_2-\theta_1^2-3)&=0
\end{aligned}
$$

Substituting $1+\theta_1^2=\theta_2$ from the second FOC, the bracketed term reduces to -4, which means $\theta_1=0$ and $\theta_2=1$. If $\mathrm{E}\left[X^3\right]=C\neq 0$, and we repeat the steps above, we end up with

$$
\begin{aligned}
&C-4\theta_1=0\\
&\theta_1=\frac{C}{4},\ \theta_2=\frac{16+C^2}{16}
\end{aligned}
$$

Question 7

The expected value of X is $\mathrm{E}[X]=\frac{A}{2}+\frac{1}{2A}=\frac{A^2+1}{2A}$. When we subtract 1, we get A^2-2A+1 or $(A-1)^2$ in the numerator, which is always positive as long as $A\neq 1$. The denominator is always positive since $A>0$, so $\mathrm{E}[X]-1>0$ and $\mathrm{E}[X]>1$.

The expected value of $\log(X)$ is $\frac{1}{2}\log(A)+\frac{1}{2}\log\left(\frac{1}{A}\right)=\frac{1}{2}\log(A)+\frac{1}{2}\log(1)-\frac{1}{2}\log(A)=0$.

Since the games are independent,

$$
\mathrm{E}[Y_n]=\prod_{i=1}^{n}\mathrm{E}[X_i]=\prod_{i=1}^{n}\frac{A^2+1}{2A}=\left(\frac{A^2+1}{2A}\right)^n
$$

which means that the expected value of the gamble is increasing in n since $\mathrm{E}[X]>1$. $\mathrm{E}[\log(Y_n)]$ is equal to the sum of $\mathrm{E}[\log(X_i)]$ for i from 1 to n; each term in the sum is equal to 0.

To show the probability convergence of Y_n, we note that

$$
\log(Y_n)=\sum_{i=1}^{n}\log(X_i)
$$

We let $Z=\log(X)$. Z is also a random variable since any measurable function of a random variable is random. $\mathrm{E}[Z]=0$ as noted before and

$$
\mathrm{Var}[Z]=\mathrm{E}[Z^2]=\frac{1}{2}[\log(A)]^2+\frac{1}{2}\left[\log\left(\frac{1}{A}\right)\right]^2=\frac{1}{2}\left\{[\log(A)]^2+[\log(1)-\log(A)]^2\right\}=[\log(A)]^2
$$

Using the Lindeberg-Lévy central limit theorem, we have

$$
\frac{1}{\sqrt{n}}\log(Y_n)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i\xrightarrow{D}N(0,[\log(A)]^2)
$$

Since $\log(\cdot)$ is a monotonic transformation,

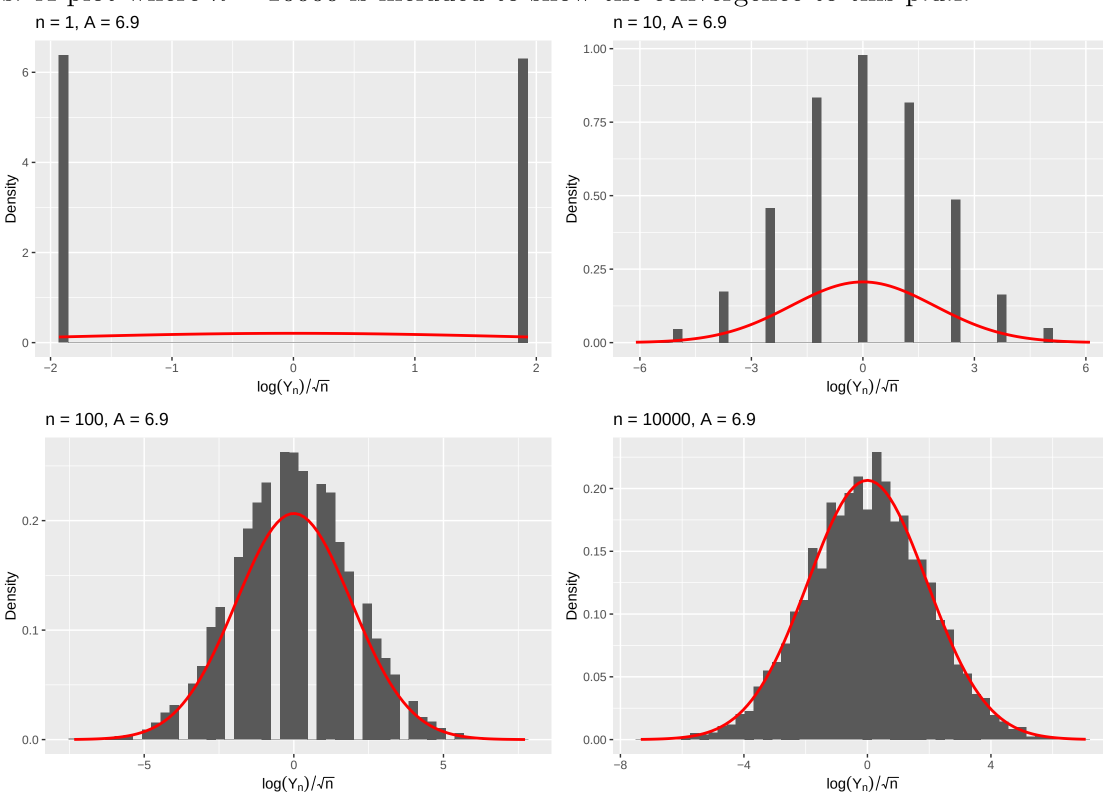
$$
\Pr(Y_n<\varepsilon)=\Pr[\log(Y_n)<\log(\varepsilon)]=\Pr\left[\frac{1}{\sqrt{n}}\log(Y_n)<\frac{1}{\sqrt{n}}\log(\varepsilon)\right]
$$

and we've seen that as n tends to infinity, this is equal to the cumulative distribution function of $N(0, [\log(A)]^2)$ evaluated at $z = \frac{1}{\sqrt{n}} \log(\varepsilon)$. But as n tends to infinity, z tends to 0, and the probability that $Z < 0$ is $\frac{1}{2}$ since it is normally distributed and centered around 0. A similar argument applies for $\Pr(Y_n < \frac{1}{\varepsilon})$ since

$$\Pr\left(Y_n < \frac{1}{\varepsilon}\right) = \Pr\left[\log(Y_n) < \log\left(\frac{1}{\varepsilon}\right)\right] = \Pr\left[\frac{1}{\sqrt{n}} \log(Y_n) < -\frac{1}{\sqrt{n}} \log(\varepsilon)\right]$$

A risk-neutral individual would be willing to pay $\mathrm{E}[Y_n]$ to play the game, although this value increases without bound as n increases. If the individual were risk-averse with a logarithmic utility function, then the expected utility of the gamble is 0 since $\mathrm{E}[\log(Y_n)] = 0$. and they are indifferent between playing and not playing unless they have to pay for the 'privilege'.

Below, the distribution of payoffs after n periods with $n = 1, 10, 100$ is shown by simulating the game 10,000 times and plotting the result. The log of Y_n is plotted instead of Y_n because Y_n is too far apart for $n = 10, 100$ to meaningfully fit on the x-axis (and there is nothing visually interesting about the distribution of the absolute payoffs). The values are also normalized by $1/\sqrt{n}$ so that we can see how the distribution compares to the p.d.f. of $N(0, [\log(A)]^2)$ which is overlaid in the plots. A plot where $n = 10000$ is included to show the convergence to this p.d.f.



Question 8

First we brute-force the question for the two-dice game. We take the son's throw to be S and the father's throw to be F. The probability that the son wins in the two-dice game is

$$\begin{aligned}\Pr(S_1+S_2>F) &= \sum_{i=1}^{6}\frac{1}{6}\Pr(S_1+i>F)\\ &= \sum_{j=1}^{6}\sum_{i=1}^{6}\frac{1}{36}\Pr(S_1+i>j)\\ &= \frac{1}{36}\sum_{j=1}^{6}\sum_{i=1}^{6}\Pr(S_1>j-i)\end{aligned}$$

We only have to consider cases where $j-i\leq 5$:

$j-i$	Possible combinations (j,i)	Number of instances	$\Pr(S_1>j-i)$
-5	$(1,6)$	1	1
-4	$(1,5),(2,6)$	2	1
-3	$(1,4),(2,5),(3,6)$	3	1
-2	$(1,3),(2,4),(3,5),(4,6)$	4	1
-1	$(1,2),(2,3),(3,4),(4,5),(5,6)$	5	1
0	$(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)$	6	1
1	$(2,1),(3,2),(4,3),(5,4),(6,5)$	5	$\frac{5}{6}$
2	$(3,1),(4,2),(5,3),(6,4)$	4	$\frac{4}{6}$
3	$(4,1),(5,2),(6,3)$	3	$\frac{3}{6}$
4	$(5,1),(6,2)$	2	$\frac{2}{6}$
5	$(6,1)$	1	$\frac{1}{6}$

And so the probability the son wins is

$$\begin{aligned}&\frac{1}{36}\left[5\times\frac{5}{6}+4\times\frac{4}{6}+3\times\frac{3}{6}+2\times\frac{2}{6}+\frac{1}{6}+(1+2+3+4+5+6)\right]\\ =&\frac{1}{36}\left[\frac{55}{6}+21\right]=\frac{181}{216}\end{aligned}$$

When we have n and $2n$ dice, we can abuse the Central Limit Theorem. For the outcome of a dice roll X_i, we have

$$\mathrm{E}[X_i]=3.5,\ \mathrm{Var}[X_i]=\mathrm{E}[X_i^2]-\mathrm{E}[X_i]^2=\frac{1}{6}[1^2+2^2+3^2+4^2+5^2+6^2]-3.5^2=\frac{35}{12}$$

When the father throws n dice, his score is $F=\sum_{i=1}^{n}X_i$ and when the son throws $2n$ dice, his score is $S=\sum_{i=1}^{2n}X_i$. We know that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-3.5)&\xrightarrow{d}N\left(0,\frac{35}{12}\right)\\ \frac{1}{\sqrt{2n}}\sum_{i=1}^{2n}(X_i-3.5)&\xrightarrow{d}N\left(0,\frac{35}{12}\right)\end{aligned}$$

We now make the assumption that n is large enough for the above to be a good approximation. After some manipulations, we have

$$F \sim N\left(3.5n, \frac{35}{12}n\right)$$
$$S \sim N\left(7n, \frac{35}{6}n\right)$$

The p.d.f. for F and S are

$$p_F(f) = \frac{1}{\sqrt{2\pi(35/12)n}} e^{-\frac{(f-3.5n)^2}{2(35/12)n}} = \frac{1}{\sqrt{\pi(35/6)n}} e^{-\frac{(f-3.5n)^2}{(35/6)n}}$$
$$p_S(s) = \frac{1}{\sqrt{2\pi(35/6)n}} e^{-\frac{(s-7n)^2}{2(35/6)n}} = \frac{1}{\sqrt{\pi(35/3)n}} e^{-\frac{(s-7n)^2}{(35/3)n}}$$

F and S are independent, so their joint probability density function is the product of their marginal p.d.f.s, and the probability that $S > F$ is

$$\begin{aligned}\int_{-\infty}^{\infty} \int_{f}^{\infty} p_{F,S}(f,s)\mathrm{d}s\mathrm{d}f &= \int_{-\infty}^{\infty} \int_{f}^{\infty} \frac{1}{\sqrt{\pi(35/6)n}} e^{-\frac{(f-3.5n)^2}{(35/6)n}} \frac{1}{\sqrt{\pi(35/3)n}} e^{-\frac{(s-7n)^2}{(35/3)n}} \mathrm{d}s\mathrm{d}f \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi(35/6)n}} e^{-\frac{(f-3.5n)^2}{(35/6)n}} \int_{f}^{\infty} \frac{1}{\sqrt{\pi(35/3)n}} e^{-\frac{(s-7n)^2}{(35/3)n}} \mathrm{d}s\mathrm{d}f\end{aligned}$$

whatever that is.