Parts

seen ⇓

(i) (e). Using Bayes Theorem,

$$\begin{aligned} P(\overline{B}|A) &= \frac{P(A \cap \overline{B})}{P(A)} = \frac{P(A|\overline{B})P(\overline{B})}{P(A)} = \frac{(1 - P(\overline{A}|\overline{B}))(1 - P(B))}{P(A)} \\ &= \frac{(1-0.8)(1-0.3)}{0.2} = 0.7. \end{aligned}$$

unseen ⇓

(ii) (d). From the standard normal table, $\Phi(1.282) \approx 0.9$ and so $\Phi(-1.282) \approx 0.1$. It follows that

$$\frac{90-100}{\sigma} \approx -1.282 \implies \sigma \approx \frac{10}{1.282} \approx 7.8.$$

seen ⇓

(iii) (e).

(iv) (f). For $X_i \sim \text{Exponential}(\lambda)$, $f(x_i) = \lambda \exp(-\lambda x_i)$. So,

$$\begin{aligned} \ell(\lambda) &= \sum_{i=1}^{3} \log f(x_i) = 3\log(\lambda) - \lambda \sum_{i=1}^{3} x_i \\ \frac{d\ell(\lambda)}{d\lambda} &= \frac{3}{\lambda} - \sum_{i=1}^{3} x_i \\ \implies \hat{\lambda} &= \frac{3}{\sum_{i=1}^{3} x_i} = \frac{3}{0.1+0.5+0.9} = \frac{3}{1.5} = 2. \end{aligned}$$

(v) (b).

Observed:

	1.9GHz	2.1GHz	
SATA	26	27	53
SSD	48	81	129
	74	108	182

Expected:

	1.9GHz	2.1GHz
SATA	21.55	31.45
SSD	52.45	76.55

Test statistic X^2=2.1853.

Each 4 marks

Question		Marks & seen/unseen
2.		

Parts

(i)	(a) $X \sim$ Geometric(0.2). From the formula sheet, $E(X) = 1/0.2 = 5$.	seen ⇓ 4 marks
	(b) $$\mathrm{P}(X \leq 2) = \mathrm{P}(X = 1) + p(X = 2) = 0.2 + 0.2 * (1 - 0.2) = 0.36.$$	3 marks
	(c) [A one-sided hypothesis test for a geometric parameter higher than 0.2 would consider the left tail of the distribution, so 0.36 is in fact the p-value for this test.] There is very little evidence to reject the null hypothesis of 80% next day deliveries.	unseen ⇓ 1 marks
(ii)	(a) $Y \sim$ Binomial$(20, 0.2)$. From the formula sheet, $E(Y) = 20 \times 0.2 = 4$.	seen ⇓ 4 marks
	(b) $$\mathrm{P}(Y \leq 5) = \sum_{i=0}^{5} \mathrm{P}(Y = i) = \sum_{i=0}^{5} \binom{20}{i} 0.2^i \times 0.8^{20-i} \approx 0.804.$$	3 marks
	(c) [A one-sided hypothesis test for a binomial probability parameter higher than 0.2 would consider the right tail of the distribution. The p-value for the test is:] $$\mathrm{P}(Y \geq 6) = 1 - \mathrm{P}(Y \leq 5) \approx 0.196.$$ So there is more evidence against the null hypothesis than before, but insufficient to reject at any typical significance level.	unseen ⇓ 1 marks
(iii)	Let Z be the number of parcels delivered on time, then $Z \sim$ Binomial$(1000, 0.8)$. So $\mathrm{E}(Z) = 1000 \times 0.8 = 800$ and $\mathrm{Var}(Z) = 1000 \times 0.8 \times (1 - 0.8) = 160$, and hence by the CLT approximately $Z \sim \mathrm{N}(800, 160)$. $P(Z \leq 774) \approx \Phi\left(\frac{774-800}{\sqrt{160}}\right) \approx \Phi(-2.055) \approx 0.020$. So with this larger sample there is now much stronger evidence that 80% is not being achieved.	seen ⇓ 4 marks

Question 3.

Marks & seen/unseen

Parts

seen ⇓

(i) Properties of a pdf:

I. $f(x) \geq 0, \forall x \in \mathbb{R}$;

II. $\int_{x=-\infty}^{\infty} f(x)dx = 1$.

unseen ⇓

From the second point, we have

$$1 = \int_{x=-\infty}^{\infty} f(x)dx = \int_{x=1}^{\infty} \frac{c}{x^3}dx = \left.\frac{-c}{2x^2}\right|_{x=1}^{\infty} = \frac{c}{2}.$$

So $c = 2$. 5 marks

(ii)

$$\mathrm{E}(X) = \int_{x=-\infty}^{\infty} xf(x)dx = \int_{x=1}^{\infty} \frac{2}{x^2}dx = \left.\frac{-2}{x}\right|_{x=1}^{\infty} = 2.$$

5 marks

(iii) To find the median, we first work out the cdf:

$$F(x) = \int_{u=-\infty}^{x} f(u)du = \int_{u=1}^{x} \frac{2}{u^3}du = \left.\frac{-1}{u^2}\right|_{u=1}^{x} = 1 - \frac{1}{x^2}.$$

The median is then x satisfying $F(x) = 1/2$.

$$\frac{1}{2} = 1 - \frac{1}{x^2} \iff \frac{1}{x^2} = \frac{1}{2} \implies x = \sqrt{2}.$$

So the median is $\sqrt{2}$. 7 marks

(iv) We have seen the median is less than the mean, suggestive of right skew. This is consistent with the shape of the density function $f(x)$, which is decreasing in x on $[1, \infty)$. 3 marks

Question 4. | Marks & seen/unseen

Parts

(i) (a) $\text{bias}(T) = \mathrm{E}[T|\theta] - \theta$. — seen ⇓ — 1 marks

(b) Obvious unbiased estimators for mean and variance are $\bar{x} = 1001.625$ and $s^2_{n-1} = 117.982$ respectively. — 6 marks

(c) Let X be the weight of a randomly chosen bag of potatoes, so approximately $X \sim \mathrm{N}(1001.625, 117.982)$. Then $\mathrm{P}(X \leq 990) \approx \Phi\left(\dfrac{990 - 1001.625}{\sqrt{117.982}}\right) \approx \Phi(-1.070) \approx 0.14$. — unseen ⇓ — 3 marks

(ii) (a) Let $z = \dfrac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} = \dfrac{1001.625 - 1000}{\sqrt{100/8}} = 0.460$. From the formula sheet, the rejection region of a two-sided normal test at the 1% level is (-2.576,2.576), so there is insufficient to reject the null hypothesis of the mean being 1kg. — seen ⇓ — 7 marks

(b) If each bag weight X_i is i.i.d. $\mathrm{N}(1000, 100)$ then the sum S_5 of the weights of five independent bags is $\mathrm{N}(5000, 500)$.

So $\mathrm{P}(S_5 > 4990) = 1 - \Phi\left(\dfrac{4990 - 5000}{\sqrt{500}}\right) \approx 1 - \Phi(-0.447) \approx 0.673$. — 3 marks