10/10

1. (Detecting Defects)

(a)

Proof. Let D to denote the event that at least one defective cookie is found. Then

$$\begin{aligned} Pr[D] &= 1-(1-p)^k && (1)\\ &\geq 1-(1-\alpha)^k && (2)\\ &\geq 1-(1-\alpha)^{\frac{\ln 100}{\alpha}} && (3)\\ &\geq 1-(e^{-\alpha})^{\frac{\ln 100}{\alpha}} && (4)\\ &\geq 1-e^{\ln 1/100} && (5)\\ &\geq 1-0.01 = 0.99 && (6)\end{aligned}$$

5/5.

Note that from (1) to (2), we have used the inequality $p \geq \alpha$; from (2) to (3), we have used the inequality $k \geq \frac{\ln 100}{\alpha}$; and from (3) to (4), we have used the inequality $1-\alpha \leq e^{-\alpha}$.

(b) Solution: Let p_i be the probability that worker i will have a defect, D be the event that if all unreliable workers are caught and let F_i be the event that the test fails to catch worker i.

$$\begin{aligned} Pr[\cup_{i=1}^n F_i] &\leq \sum_{i=1}^n Pr[F_i] && (7)\\ Pr[D] &= 1-Pr[\cup_{i=1}^n F_i] && (8)\\ &\geq 1-\sum_{i=1}^n Pr[F_i] && (9)\\ &= 1-\sum_{i=1}^n (1-p_i)^k && (10)\\ &\geq 1-n(1-\alpha)^k && (11)\end{aligned}$$

From (10) to (11), we've used the fact that $p_i \geq \alpha$ if worker i is unreliable. In order for $Pr[D]$ to be at least 99%, then

$$1-n(1-\alpha)^k \geq 99\%$$

Solving for k,

$$\begin{aligned} 1-n(1-\alpha)^k &\geq 0.99 && (12)\\ n(1-\alpha)^k &\leq 0.01 && (13)\\ (1-\alpha)^k &\leq 0.01/n && (14)\\ k &\geq \log_{1-\alpha}(0.01/n) && (15)\end{aligned}$$

5/5.

7/10

2. (Improved Randomized Min-Cut Algorithm)

(a) Solutions: If we run the algorithm twice, then the number of edge contractions will be two times the number of contractions in a single run, $2(n-2)$. ✓

The probability of missing a min cut in both runs will be bounded by

$$\left(1-\frac{2}{n(n-1)}\right)^2$$

Thus, the probability of finding a min cut with two runs will be bounded by

$$1-\left(1-\frac{2}{n(n-1)}\right)^2$$ ✓

2/2

(b) Solutions: When we contract the graph from n vertices down to k vertices, the number of contraction is $n-k$. For the second step, the number of contractions is $l(k-2)$. So the total number of contraction is $(n-k)+l(k-2)$.

Assume that the number of edges in the min-cut set C is m. Let E_i be the event that the edge contracton contracted in iteration i is not in C, adn let $F_i = \cap_{j+1}^{i} E_j$. In the first contraction, the probability of not choosing an edge in the min-cut set C to contract

$$Pr[E_1] = Pr[F_1] \geq 1-\frac{2m}{nm} = 1-\frac{2}{n}.$$

Conditioned on the event that in the first contraction, the edge contracted is not in C, we can continue the analysis,

$$Pr[E_2|F_1] \geq 1-\frac{m}{m(n-1)/2} = 1-\frac{2}{n-1}.$$

Similarly,

$$Pr[E_i|F_{i-1}] \geq 1-\frac{m}{m(n-i+1)/2} = 1-\frac{2}{n-i+1}.$$

Thus, the probability that the edges in the set C have never been contracted until we are left with k vertices is

$$\begin{aligned}Pr[F_{n-k}] &= Pr[E_{n-k}|F_{n-k-1}] \cdot Pr[E_{n-k-1}|F_{n-k-2}] \cdots Pr[E_2|F_1] \cdot Pr[F_1] \\ &\geq \prod_{i=1}^{n-k}\left(1-\frac{2}{n-i+1}\right) = \prod_{i=1}^{n-k}\left(\frac{n-i-1}{n-i+1}\right) \quad ✓ \\ &= \frac{k(k-1)}{n(n-1)} \quad ✓\end{aligned}$$

Similarly for the second stage, for each copy of the l problems, we can apply the same analysis and bound the probability of finding the min-cut set of size m by

$$\frac{2}{k(k-1)}$$

4/4

Thus, the totally probability of finding the min-cut set after the two steps is bounded by

$$\frac{k(k-1)}{n(n-1)}\left(1-(1-\frac{2}{k(k-1)})^l\right)$$ ✓

(c) Solutions: The problem can be formulated as the following optimization problem,

$$\max_{k,l} \frac{k(k-1)}{n(n-1)}\left(1-(1-\frac{2}{k(k-1)})^l\right)$$
$$\text{s.t. } (n-k)+l(k-2)=2(n-2)$$

Also n, k, l have their range constraints

5/4

8/10

3. (Jensen's Inequality)

3/3

(a) Solutions: If f is a concave function, then

$$\mathbb{E}(f(X)) \leq f(\mathbb{E}(X)) \tag{1}$$

This follows naturally from the original Jensen's inequality because if f is concave, then $-f$ will be convex, and we can apply the Jensen's equality to $-f$, we would obtain the equality for concave function f as above.

(b)

3/4

Proof. Let $G_n = \sqrt[n]{\prod_{i=1}^{n} x_i}$ be the geometric mean and $A_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ be the arithmetic mean of a collection of n positive real numbers $\{x_i\}$.

$$\begin{aligned}
\log A_n &= \log\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) && (2)\\
&\geq \frac{1}{n}\sum_{i=1}^{n}\log x_i && (3)\\
&= \sum_{i=1}^{n}\left(\log x_i^{1/n}\right) && (4)\\
&= \log\left(\prod_{i=1}^{n} x_i^{1/n}\right) && (5)\\
&= \log G_n && (6)\\
A_n &\geq G_n && (7)
\end{aligned}$$

(a) says a statement on a r.v. what is the r.v. and with what prob?

needs monotonicity

From (2) to (3) we have used the inequality in (a) where $f(x) = \log x$. (7) is obtained when we take the exponential of $\log A_n$ and $\log G_n$ respectively.

(c)

2/3

Proof. let $f(x) = \sin x$ where $0 < x < \pi$. Because $f''(x) = -\sin x < 0$ when $0 < x < \pi$, $f(x)$ is concave in the interval $(0, \pi)$. We can apply the inequality in (a),

$A, B, C \in [0, 180^\circ)$ Again, what are the r.v.s?

$$\begin{aligned}
\frac{1}{3}\Big(\sin A + \sin B + \sin C\Big) &\leq \sin\frac{1}{3}(A + B + C) && (8)\\
&\leq \sin 60^\circ && (9)\\
&\leq \frac{\sqrt{3}}{2} && (10)\\
\sin A + \sin B + \sin C &\leq \frac{3\sqrt{3}}{2} && (11)
\end{aligned}$$

10/10

4. (Random Children)

(a) Solution: Let C, B, G denote the RVs the total number of children, the total number of boys, and the total number of girls the couple will have respectively. Then $C \sim Geom(1/2)$, and $\mathbb{E}(C) = 1/(1/2) = 2$. Since the couple will end until they have a girl, thus, $\mathbb{E}(G) = 1$. $\mathbb{E}(B) = \mathbb{E}(C) - \mathbb{E}(G) = 2 - 1 = 1$. Thus, both the expected number of girls and boys are 1.

(b) Solution: If the probability of having a girl is only 0.4. Then $C \sim Geom(0.4)$, and $\mathbb{E}(C) = 1/(0.4) = 2.5$, and $\mathbb{E}(G) = 1$. Then $\mathbb{E}(B) = \mathbb{E}(C) - \mathbb{E}(G) = 2.5 - 1 = 1.5$.

(c) Solution: Let C_1, B_1, G_1 denote the RVs the total number of children, the total number of boys, and the total number of girls the couple will have respectively following the new rule and p the probability of having a girl.

We can construct other variables C_2, B_2 and G_2 which denote the additional children the couple would have if they follow the old rule.

$$
\begin{align}
C &= C_1 + C_2 \tag{1}\\
B &= B_1 + B_2 \tag{2}\\
G &= G_1 + G_2 \tag{3}\\
C_1 &= B_1 + G_1 \tag{4}\\
C_2 &= B_2 + G_2 \tag{5}\\
C &= B + G \tag{6}
\end{align}
$$

Note that $C_2 \sim Geom(p)$ happens with probability $(1-p)^k$ which is denote by $Pr[C > k]$ otherwise $C_2 = 0$.

$$
\begin{align}
\mathbb{E}(B_1) &= \mathbb{E}(B) - Pr[C > k]\mathbb{E}(B_2) \nonumber\\
&= (1/p - 1) - (1-p)^k(1/p - 1) \nonumber\\
&= \left[1 - (1-p)^k\right](1/p - 1) \tag{7}
\end{align}
$$

Similarly, for the expected number of girls, G_1, under the new rule,

$$
\begin{align}
\mathbb{E}(G_1) &= \mathbb{E}(G) - Pr[C > k]\mathbb{E}(G_2) \nonumber\\
&= 1 - (1-p)^k 1 \nonumber\\
&= 1 - (1-p)^k \tag{8}
\end{align}
$$

If we plug in the value $p = 0.5$ to (7) and (8) respectively, we obtain the expected number of boys to be $1 - (1/2)^k$, and the expected number of girls to be $1 - (1/2)^k$.

(d) We plug in the value $p = 0.4$ to (7) and (8) respectively, we obtain the expected number of boys to be $1.5(1 - 0.6^k)$, and the expected number of girls to be $1 - 0.6^k$.