Homework 4 solutions

1

(a)

Variance is defined as $Var(X) = E(X^2) - E(X)^2$. Therefore $Var(X+Y) = E((X+Y)^2) - E(X+Y)^2 = E(X^2+2XY+Y^2) - (E(X)+E(Y))^2 = E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 + 2E(XY) - 2E(X)E(Y)$, due to the linearity of expectation. If X and Y are independent, $E(XY) = E(X)E(Y)$, and the expression simplifies to $E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 = Var(X) + Var(Y)$.

(b)

The above proof does not hold if X and Y are not independent. This is because the proof assumes that $E(XY) = E(X)E(Y)$, which may not be the case if X and Y are not independent. In that case, $Var(X+Y) = Var(X) + Var(Y) + 2(E(XY) - E(X)E(Y))$.

2

(a)

$\frac{M}{\min} > 6d \iff \min < \frac{M}{6d}$. This justifies the first equality, since $\mathbb{P}(A) = \mathbb{P}(B)$ if $A \iff B$. The second equality comes from the fact that all elements in S are of the form $h(b_k)$ for some k, and $textmin$ is the smallest element in S, so $\min = h(b_k)$ for some k. Thus, the statement $\min < \frac{M}{6d}$ is equivalent to the statement $\exists k : h(b_k) < \frac{M}{6d}$. Finally, the event $A = \exists k : h(b_k) < \frac{M}{6d}$ is a subset of the event $B = \bigcup_{i=1}^{|S|} h(b_i) < \frac{M}{6d}$, and therefore $\mathbb{P}(A) \leq \mathbb{P}(B)$. $\mathbb{P}(B) = \mathbb{P}\left(\bigcup_{i=1}^{d} h(b_i) < \frac{M}{6d}\right) = \sum i = 1^d \mathbb{P}\left(h(b_i) < \frac{M}{6d}\right)$, by the third axiom of probability.

(b)

(c)

$\mathbb{P}(y=0) = \mathbb{P}\left(\left(\sum i = 1^d y_i\right) = 0\right) = \mathbb{P}\left(\bigcup_{i=1}^{d}(y_i = 0)\right)$, which, by definition, is equal to $\mathbb{P}\left(\bigcup_{i=1}^{d}\left(h(b_i) > \frac{6M}{d}\right)\right)$. $h(b_k) < \frac{6M}{d}$ is the event that the hash value is *not* in $\left[0, \frac{6M}{d}\right]$, so $\bigcup_{i=1}^{d}\left(h(b_i) > \frac{6M}{d}\right)$ is the event that none of the hash values are in this range. The probability of at least one hash value being in this range is, therefore, the probability of this event not happening. Therefore, saying $\mathbb{P}(y > 0)$ is high, and therefore $\mathbb{P}(y = 0)$ is low, is equivalent to saying that the probability of seeing at least one hash value in this range is high.

(d)

(e)

By the definition of expectation, $E(y_i) = \sum n \cdot \mathbb{P}(y_i = n) \geq 1 \cdot \mathbb{P}(y_i = 1)$. Since $\mathbb{P}(y_i = 1) \geq \frac{6}{d}$, we have $E(y_i) \geq \frac{6}{d}$. By definition, $E(y) = \sum_{i=1}^{d} E(y_i) \geq d \cdot \frac{6}{d} = 6$.

(f)

By definition, $Var(y_i) = E(y_i^2) - E(y_i)^2$. Since the only values y_i can take are 0 and 1, $y_i^2 = y_i$ since $0^2 = 0$ and $1^2 = 1$. Therefore, $Var(y_i) = E(y_i) - E(y_i)^2$. Since y_i only takes positive values, $E(y_i)$ must also be positive. Therefore, $Var(y_i) = E(y_i) - E(y_i)^2 \leq E(y_i)$.

(g)

If $y = 0$, then $|y - E(y)| = E(y) \geq E(y)$. However, the reverse is not true. Therefore, the event that $y = 0$ is a subset of the event that $|y - E(y)| \geq E(y)$, and so $\mathbb{P}(y = 0) \leq \mathbb{P}(|y - E(y)| \geq E(y))$.

(h)

This is simply a consequence of Chebyshev's inequality, which states that for any random variable X and number a, $\mathbb{P}(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}$. The result in this step comes from substituting $X = y$ and $a = E(y)$.

3

(a)

To prove that a function is a bijection, it suffices to show that it is both injective and surjective.

Let A and B be two events such that $I(A) = I(B)$, i.e. their indicator variables are equal. The indicator variable is, by definition, 1 is the event is true and 0 if the event is false. If $I(A) = I(B)$, then their indicator variables always take on the same value as each other, that is, $A \iff B$. This means that the two events are the same, that is, $A = B$. This means that I is injective.

Let a be some indicator variable. By definition, a is 1 if a certain event in S is true, and 0 if it is false. Let that event be A. Then, $I(A) = a$. Since no assumptions have been made about a, this means that for any $a \in [S \to \{0, 1\}]$, there exists $A \in S$ such that $I(A) = a$, and therefore I is surjective.

Since I is both surjective and injective, it is bijective.

(b)

Let X be a random variable that is uniformly distributed over the set $\{-1, 1\}$, and let $Y = -X$. X and Y are different, but they have the same PMF. Therefore, $PMF(X) = PMF(Y)$, while $X \neq Y$. Therefore the function is not injective.

Let $f : \mathbb{R} \to \mathbb{R}$ be the function mapping $x \mapsto 2$. Since the value of a PMF can never exceed 1 (due to the first axiom of probability), there is no $X \in [S \to \mathbb{R}]$ such that $PMF(X) = f$. Therefore, PMF is not surjective.