Homework 4

Asher Christian 006-150-286

2024-05-12

1 Problem 1

Suppose we roll a die, let X be the number we get. Suppose the probability mass function $p(x) = P(X = x)$ is such that $p(1) = .1$, $p(2) = .1$, $p(3) = .1$, $p(4) = .2$, $p(5) = .2$, $p(6) = .3$

1. Calculate $P(X > 4)$. Calculate $P(X = 6|X > 4)$
   $P(X > 4) = p(5) + p(6) = 0.2 + 0.3 = 0.5$
   $P(X = 6|X > 4) = \frac{P(X=6 \cap X>4)}{P(x>4)} = \frac{p(6)}{p(5)+p(6)} = 0.6$

2. Calculate $E(X)$, $Var(X)$, and $SD(X)$

   $$E(X) = \sum_{i=1}^{6} ip(i) = 1*0.1+2*0.1+3*0.1+4*0.2+5*0.2+6*0.3 = 4.2.$$

   $$Var(X) = E((X - E(X))^2) =$$

   $(1-4.2)^2*0.1+(2-4.2)^2*0.1+(3-4.2)^2*0.1+(4-4.2)^2*0.2+(5-4.2)^2*0.2+(6-4.2)^2*0.3 = 2.76.$

   $Var(x) = E(X^2)-E(X)^2 = 1*0.1+4*0.1+9*0.1+16*0.2+25*0.2+36*0.3-17.64 = 2.76.$

   $SD(X) = \sqrt{Var(x)} \approx 1.67$

3. Suppose the reward for x is $h(x)$, and $h(1) = -\$20$, $h(2) = -\$10$, $h(3) = \$0$, $h(4) = \$10$, $h(5) = \$20$, $h(6) = \$100$. Calculate $E(h(X))$, $Var(h(X))$ and $SD(h(X))$. What are the units of $E(h(X))$ and $Var(h(X))$

   $$E(h(x)) = \sum_{i=1}^{6} h(i)p(i) = -20*0.1-10*0.1+0*0.1+10*0.2+20*0.2+100*0.3 = 33.0\$.$$

   $$\begin{aligned} Var(h(x)) &= E(h(x)^2) - E(h(x))^2 \\ &= 400*0.1 + 100*0.1 + 0*0.1 + 100*0.2 + 400*0.2 + 10000*0.3 - 1089 \\ &= 2061\$^2 \end{aligned}$$

   $$SD(h(x)) = \sqrt{Var(h(x))} \approx 45.40\$.$$

   The units for $E(X)$ are \$ and the units for $Var(X)$ are $\$^2$

2 Problem 2

Suppose $Z \in \{0, 1\}$. $P(Z = 1) = p$, $P(Z = 0) = 1 - p$. Calculate $E(Z)$, $E(Z^2)$ and $Var(Z)$. What if we replace 0 by -1? Calculate concrete numbers for $p = 1/2$.

$$E(Z) = \{p = 0.5, p - (1 - p) = 0\}.$$

$$E(Z^2) = \{p = 0.5, p + (1 - p) = 1\}.$$

$$Var(Z) = E(Z^2) - E(Z)^2 = \{p - p^2 = 0.25, (p + (1 - p)) - (p - (1 - p))^2 = 1\}.$$

3 Problem 3

Suppose we flip a fair coin 100 times independently. Let X be the number of heads. Calculate $E(X)$, $Var(X)$, $SD(X)$, $E(X/100)$, $Var(X/100)$, $SD(X/100)$. Write down the formula for computing $P(X \in [40, 60])$

$$E(X) = \sum_{i=0}^{100} i \frac{\binom{100}{i}}{2^{100}} = 50.$$

$$Var(X) = E(X^2) - E(X)^2 = \sum_{i=0}^{100} i^2 \frac{\binom{100}{i}}{2^{100}} - 2500 = 25.$$

$$SD(X) = \sqrt{Var(X)} = 5.$$

$$E(\frac{X}{100}) = \sum_{i=0}^{100} \frac{i}{100} \frac{\binom{100}{i}}{2^{100}} = 0.5.$$

$$Var(\frac{X}{100}) = E((\frac{X}{100})^2) - E(\frac{X}{100})^2 = \sum_{i=0}^{100} \frac{i^2}{10000} \frac{\binom{100}{i}}{2^{100}} - 0.5 = 0.0025.$$

$$SD(\frac{X}{100}) = \sqrt{Var(X)} = 0.05.$$

$$P(X \in [40, 60]) = \sum_{i=40}^{60} \frac{\binom{100}{i}}{2^{100}} \approx 0.964799799782.$$

4 Problem 4

Suppose within the population of voters, 20% of them support a candidate A. If we randomly sample 100 people sequentially with replacement. Let X be the number of supporters of A among these 100 people. Then what is the distribution of X? What are $E(X)$, $Var(X)$, and $SD(X)$? What are $E(X/100)$, $Var(X/100)$, and $SD(X/100)$?
X could be anywhere in the range [0,100] though it will be most likely $E(X)$ and taper off with standard distribution $SD(X)$

Let 1 be the event that a individual supports candidate A and 0 the event that the individual does not support A

$$E(X) = \sum_{i=0}^{100} i \binom{100}{i} (0.20)^i (0.80)^{100-i} = 20.$$

$$Var(X) = E(X^2) - E(X)^2 = \sum_{i=0}^{100} i^2 \binom{100}{i} (0.20)^i (0.80)^{100-i} - 400 = 16.$$

$$SD(X) = \sqrt{Var(X)} = 4.$$

$$E(\frac{X}{100}) = \frac{E(X)}{100} = 0.2.$$

$$Var(\frac{X}{100}) = E(\frac{X^2}{100^2}) - (\frac{E(X)}{100})^2 = \frac{E(X^2)}{100^2} - \frac{E(X)^2}{100^2} = \frac{Var(X)}{100^2} = 0.0016.$$

$$SD(\frac{X}{100}) = \sqrt{Var(\frac{X}{100})} = 0.04.$$

5 Problem 5

Suppose we randomly throw 10,000 points into the unit square $[0,1]^2$. Let A be the region $x^2 + y^2 \leq 1$. Let m be the number of points that fall into A. What is the distribution of m? Let $\hat{\pi} = 4m/10000$ be our Monte Carlo estimate of pi. What are $E(\hat{\pi})$, $Var(\hat{\pi})$ and $SD(\hat{\pi})$?

$$E(\hat{\pi}) = E(m) * \frac{4}{10000} = \frac{4np}{10,000} = \frac{(4)(10,000)\frac{\pi}{4}}{10,000} = \pi.$$

$$Var(\hat{\pi}) = np(1-p)\frac{4^2}{10,000^2} = \pi\frac{4-\pi}{10,000} \approx 0.000269676621327.$$

$$SD(\hat{\pi}) = \sqrt{Var(\hat{\pi})} \approx 0.0164218336774.$$

6 Problem 6

Suppose X is a discrete random variable with probability mass function $p(x)$, where x takes values in a discrete set.

1. Prove $E(aX) = aE(X)$.

$$E(aX) = \sum_{i=0}^{n} aX_i p(x) = a\sum_{i=0}^{n} X_i p(x) = aE(x).$$

2.

$$E(X+b) = \sum_{i=0}^{n} (X_i+b)p(x) = \sum_{i=0}^{n} X_i p(x) + \sum_{i=0}^{n} bp(x) = E(X) + b\sum_{i=0}^{n} p(x) = E(X)+b.$$

3.

$$Var(aX) = \sum_{i=0}^{n}(aX_i - E(aX))^2 p(x) = sum_{i=0}^{n}(a(X_i - E(X)))^2 p(x) =$$

$$\sum_{i=0}^{n} a^2 (X_i - E(X))^2 p(x) = a^2 Var(X).$$

4.

$$\begin{aligned}
Var(X+b) &= \sum_{i=0}^{n}((X_i + b) - E(X+b))^2 p(x) \\
&= \sum_{i=0}^{n}(X_i + b - E(X) - b)^2 p(x) \\
&= \sum_{i=0}^{(X_i - E(X))^2 p(x)} \\
&= Var(X)
\end{aligned}$$

5.

$$\begin{aligned}
Var(X) &= \sum_{i=0}^{n}(X_i - E(X))^2 p(x) \\
&= \sum_{i=0}^{n}(X_i^2 - 2X_i E(X) + E(X)^2) p(x) \\
&= E(X^2) - 2E(X)\sum_{i=0}^{n} X_i p(x) + E(X)^2 \sum_{i=0}^{n} p(x) \\
&= E(X^2) - 2E(X)^2 + E(X)^2 \\
&- E(X^2) - E(X)^2
\end{aligned}$$

6. $\mu = E(X)$, $\sigma^2 = Var(X)$, $Z = \frac{X-\mu}{\sigma}$

$$\begin{aligned}
E(Z) &= \frac{E(X) - \mu}{\sigma} \\
&= \frac{0}{\sigma} \\
&= 0 \\
Var(Z) &= \frac{Var(X - \mu)}{\sigma^2} \\
&= \frac{Var(X)}{Var(X)} \\
&= 1
\end{aligned}$$

7 Problem 7

1. For $X \sim$ Binomial(n, p), prove formally that $E(X) = np$ and $Var(X) = np(1-p)$

$$i' = i - 1, n' = n - 1.$$

$$\begin{aligned}
E(X) &= \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=0}^{n} i \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} np \frac{(n-1)!}{(i-1)!(n-i)!} p^{i-1} (1-p)^{n-i} \\
&= np \sum_{i=0}^{n'} \frac{n'!}{i'!(n'-i')!} p^{i'} (1-p)^{n'-i'} \\
&= np
\end{aligned}$$

$$i' = i - 2, n' = n - 1.$$

$$\begin{aligned}
E(X(X-1)) &= \sum_{i=0}^{n} i(i-1) P(X = i) \\
&= \sum_{i=0}^{n} i(i-1) \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\
&= \sum_{i=2}^{n} n(n-1)p^2 \frac{(n-2)!}{(i-2)!(n-i)!} p^{i-2} (1-p)^{n-i} \\
&= n(n-1)p^2 \sum_{i'=0}^{n'} \binom{n'}{i'} p^{i'} (1-p)^{n'-i'} \\
&= n(n-1)p^2 \qquad = E(X^2) - E(X)
\end{aligned}$$

$$E(X^2) = n(n-1)p^2 + np \to E(X^2) - E(X)^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p).$$

2. For $T \sim$ Geometric(p), prove $E(T) = \frac{1}{p}$

$$q = 1 - p.$$

$$
\begin{aligned}
E(T) &= \sum_{i=0}^{\infty} ipq^{i-1} \\
&= p\sum_{i=0}^{\infty} iq^{i-1} \\
&= p\sum_{i=0}^{\infty} \frac{d}{dq}q^{i} \\
&= p\frac{d}{dq}(\frac{1}{1-q}-1) \\
&= \frac{p}{(1-q)^2} \\
&= \frac{1}{p}
\end{aligned}
$$

8 Problem 8

Read the slides on Jensen inequality. For a convex function $h(x)$, and for any random variable X, prove $h(E(X)) \leq E(h(X))$.

$$
\begin{aligned}
h(E(X)) &= h(\sum_{i=0}^{n} X_i p(i)) \\
E(h(X)) &= \sum_{i=0}^{n} h(X_i)p(i)
\end{aligned}
$$

let b be the slope of the tangent line at $x = E(X)$
By definition of convexity, $h(x_0)$ is always greater than the tangent line at x_0.

$$
\begin{aligned}
h(X) &\geq h(E(X)) + b(X - E(X)) \\
E(h(x)) &\geq E(h(E(X)) + b(X - E(X)) \\
&\geq h(E(X)) + E(b(X - E(X)) \\
&\geq h(E(X)) + bE(X) - bE(X) \\
&\geq h(E(X))
\end{aligned}
$$

9 Probelm 9

Read the slides on entropy. Explain that for a probability mass function $p(x)$, its entropy can be defined by $E[-log_2 p(X)] = -\sum_x p(x)log_2(p(x))$. Explain that entropy can be interpreted as average number of coin flips or average code length

for a probability mass function p(x), the $-log_2(p(x))$ signifies the number of binary choices that would be made to get to that point. For example if $P(A) = \frac{1}{16}$, the action of A occuring represents the same amount of information as flipping a coin 4 times and taking only one of the resulting sequences, e.g HTHH. Taking the weighted average of these $-log_2$ quantities weighted by their probability, gives a measure of how many bits of information the probability distribution represents. If you are working with a code that can take on two values then the entropy of the system determines the average code length of the elements in the probability distribution, this is because the $-log_2$ of the number represents the number of two way divisions that must take playce to get to the value.