B555 - Machine Learning - Homework 1

Enrique Areyan
February 07, 2015

Problem 1: Let $(\Omega, \mathcal{F}, P)$ be a probability space. Using only the set operations and axioms of probability, show that for any two sets, not necessarily disjoint, $A \subseteq \Omega$ and $B \subseteq \Omega$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Proof: Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A \subseteq \Omega$ and $B \subseteq \Omega$. Then we can write:

$$\begin{aligned} A \cup B &= (A \cup B) \cap \Omega && \text{since both } A \subseteq \Omega \text{ and } B \subseteq \Omega \\ &= (A \cup B) \cap (B \cup B^c) && \text{since } B \cup B^c = \Omega \\ &= B \cup (A \cap B^c) && \text{by distributivity} \end{aligned}$$

Note that B is disjoint from $A \cap B^c$ since any element in B is necessarily not in B^c and hence not in $A \cap B^c$. It follows by the axioms of probability that:

$$P(A \cup B) = P(B \cup (A \cap B^c)) = P(B) + P(A \cap B^c) \quad (*)$$

But consider the following:

$$\begin{aligned} A &= A \cap \Omega && \text{since } A \subseteq \Omega \\ &= A \cap (B \cup B^c) && \text{since } B \cup B^c = \Omega \\ &= (A \cap B) \cup (A \cap B^c) && \text{by distributivity} \end{aligned}$$

Note that $A \cap B$ is disjoint from $A \cap B^c$ since any element in $A \cap B$ is necessarily in B and hence not in B, which in turn means that is not in $A \cap B^c$.

$$P(A) = P((A \cap B) \cup (A \cap B^c)) = P(A \cap B) + P(A \cap B^c) \Longrightarrow P(A \cap B^c) = P(A) - P(A \cap B) \quad (**)$$

Finally, replacing $(**)$ into $(*)$, we obtain the result:

$$P(A \cup B) = P(B) + P(A \cap B^c) = P(B) + P(A) - P(A \cap B) \Longrightarrow \boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Problem 2: Prove the following expression or provide a counterexample if it does not hold. It follows by the axioms of probability that:

$$P(A) = P(A|B) + P(A|B^c)$$

Solution: This expression is **not true**, as the following counterexample shows: consider the experiment of rolling a fair six-sided die and the events:

$$A = \text{a 6 came up}, \qquad \text{and} \qquad B = \text{an even number came up} \Longrightarrow B^c = \text{an odd number came up}$$

Since the die is fair, it follows that $P(A) = \frac{1}{6}$. Now, let us compute $P(A|B)$ and $P(A|B^c)$ separately:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} && \text{by definition of conditional probabilities} \\ &= \frac{P(\text{a 6 came up AND an even number came up})}{P(\text{an even number came up})} && \text{by definition of the events } A \text{ and } B \\ &= \frac{P(\text{a 6 came up})}{P(\text{an even number came up})} && \text{since } A \subset B \\ &= \frac{1/6}{1/2} \\ &= \frac{2}{6} = \frac{1}{3} \end{aligned}$$

$$
\begin{aligned}
P(A|B^c) &= \frac{P(A\cap B^c)}{P(B^c)} && \text{by definition of conditional probabilities}\\
&= \frac{P(\text{a 6 came up AND an odd number came up})}{P(\text{an odd number came up})} && \text{by definition of the events } A \text{ and } B\\
&= \frac{P(\emptyset)}{P(\text{an odd number came up})} && \text{since } A\cap B=\emptyset\text{, i.e., 6 is not an odd number}\\
&= \frac{0}{1/2}\\
&= 0
\end{aligned}
$$

From this computations we can conclude that the expression is not true since:

$$P(A) = \frac{1}{6} \neq \frac{1}{3} = \frac{1}{3} + 0 = P(A|B) + P(A|B^c)$$

Problem 3: Two players perform a series of coin tosses. Player one wins a toss if the coin turns heads and player two wins if it turns tails. The game is played until one player wins n times. However, the game is interrupted when player one had m wins and player two had l wins, where $0 \leq m < n$ and $0 \leq l < n$.

a) Assuming $n = 8, m = 4$, and $l = 6$, what is the probability that one would win the game if the game was to be continued later.

b) Derive the general expression or write an algorithm that player one will win the game if the game is to be continued later. Your expression should be a function of m, l, and n. If you are providing an algorithm, implement it and submit your code along with your pseudocode. You many *not* simulate the game as your solution.

Solution: a) Let M = number of (additional) wins by player one and L = number of (additional) win by player 2. In this case, we can write this probability as follows:

$$
\begin{aligned}
P(\text{player one wins}) &= P(\text{player one has 8 heads before player two has 8 tails}) && \text{rewriting events}\\
&= P(M = 8-4 \text{ before } L = 8-6) && \text{by convention mentioned before}\\
&= P(M = 4 \text{ before } L = 2) && \text{by convention mentioned before}\\
&= P\left((M = 4 \text{ AND } L = 0)\text{OR}(M = 4 \text{ AND } L = 1)\right) && \text{partitioning the event}\\
&= P\left((M = 4 \text{ AND } L = 0)\right) + P\left((M = 4 \text{ AND } L = 1)\right) && \text{sum rule}
\end{aligned}
$$

Now we can compute each of these probabilities separately, where H indicates that a toss resulted in a Head and T in a Tail:

$$
\begin{aligned}
P(M = 4 \text{ AND } L = 0) &= P(\{HHHH\}) && \text{getting four straight heads so that player one wins}\\
&= \frac{1}{2^4} && \text{since tosses are independent and the coin is fair.}\\
&= \frac{1}{16} && \text{arithmetic}
\end{aligned}
$$

$$
\begin{aligned}
P(M = 4 \text{ AND } L = 1) &= P(\{THHHH, HTHHH, HHTHH, HHHTH\})\\
&= 4 * \frac{1}{2^5} && \text{fair coin and independent tosses}\\
&= \frac{2}{16} && \text{arithmetic}
\end{aligned}
$$

Therefore, the probability we want is:

$$P(\text{player one wins}) = P(M = 4 \text{ AND } L = 0) + P(M = 4 \text{ AND } L = 1) = \frac{1}{16} + \frac{2}{16} = \boxed{\frac{3}{16}}$$

b) Let us derive a general expression for n, m and l, assuming $0 \leq m < n$ and $0 \leq l < n$. Again, let us denote the number of wins by player one as M and the number of wins by player 2 as L.

$$\begin{aligned} P(\text{player one wins}) &= P(M = n \text{ before } L = n) && \text{equivalent events} \\ &= P\left[(M = n \text{ AND } L = l) \text{ OR } (M = n \text{ AND } L = l+1) \cdots \text{ OR } (M = n \text{ AND } L = n-1)\right] \\ &= \sum_{i=l}^{n-1} \mathrm{P}(M = n \ \text{ AND } \ L = i) && \text{sum rule} \end{aligned}$$

So we have broken down the problem into computing probabilities for the event $M = n$ AND $L = i$ for $0 \leq i < n$. Note that in this event the number of wins by player two are fixed, so we can solve this problem easily. First, we need to complete a win by player one, which implies getting $n - m$ more heads. Tosses are independent, so this probability is $\left(\frac{1}{2}\right)^{n-m}$. Next, there will be i tails, which gives a probability of $\left(\frac{1}{2}\right)^{i}$. Finally, tails can occur in i of exactly $n - m + i - 1$ different places, where $n - m$ is the different places to have Heads and complete a win, to which we must add i places for the tails and subtract one to account for the fact that the last toss must always be a head. Since order does not matter, there are $\binom{n-m+i-1}{i}$ places to put i tails. Putting all of this together, we conclude that for a given i:

$$P(M = n \text{ AND } L = i) = \binom{n-m+i-1}{i} \left(\frac{1}{2}\right)^{n-m} \left(\frac{1}{2}\right)^{i} = \binom{n-m+i-1}{i} \left(\frac{1}{2}\right)^{n-m+i}$$

The final solution is the sum of all i where $0 \leq i < n - (l+1)$, since we must account for tosses won by player 2 already:

$$\boxed{P(\text{player one wins}) = \sum_{i=0}^{n-(l+1)} P(M = n \text{ AND } L = i) = \sum_{i=0}^{n-(l+1)} \binom{n-m+i-1}{i} \left(\frac{1}{2}\right)^{n-m+i}}$$

Note that this formula works for the case in part a), i.e., if $n = 8, m = 4$ and $l = 6$, then:

$$\begin{aligned} P(\text{player one wins}) = \sum_{i=0}^{8-(6+1)} \binom{8-4+i-1}{i} \left(\frac{1}{2}\right)^{8-4+i} &= \sum_{i=0}^{1} \binom{3+i}{i} \left(\frac{1}{2}\right)^{4+i} \\ &= \binom{3+0}{0} \left(\frac{1}{2}\right)^{4+0} + \binom{3+1}{1} \left(\frac{1}{2}\right)^{4+1} \\ &= \binom{3}{0} \left(\frac{1}{2}\right)^{4} + \binom{4}{1} \left(\frac{1}{2}\right)^{5} \\ &= 1 \cdot \frac{1}{16} + 4 \cdot \frac{1}{2^5} \\ &= 1 \cdot \frac{1}{16} + \frac{2}{16} \\ &= \frac{3}{16} \end{aligned}$$

The reader can also verify, using the Binomial Theorem, that in the case that $m = l$, the probability of player one winning is $1/2$, as it should since in this case both players restart the game with the same advantage.

Problem 4: Let $\Omega_X = \{a, b, c\}$ and $p_X(a) = 0.1, p_X(b) = 0.2$, and $p_X(c) = 0.7$. Let

$$f(x) = \begin{cases} 10 & x = a \\ 5 & x = b \\ 10/7 & x = c \end{cases}$$

a) What is $E[f(x)]$?

b) What is $E[1/p_X(x)]$?

c) For an arbitrary pmf $p_X(x)$, what is $E[1/f(x)]$?

Solution: a)

$$E[f(x)] = \sum_{x\in\Omega_X} f(x)p_X(x) = f(a)p_X(a)+f(b)p_X(b)+f(c)p_X(c) = 10\cdot 0.1+5\cdot 0.2+10/7\cdot 0.7 = 1+1+1 = \boxed{3}$$

b)

$$\begin{aligned} E\left[\frac{1}{p_X(x)}\right] &= \sum_{x\in\Omega_X} \frac{1}{p_X(x)} p_X(x) \\ &= \frac{1}{p_X(a)}p_X(a) + \frac{1}{p_X(b)}p_X(b) + \frac{1}{p_X(c)}p_X(c) \\ &= = 1 + 1 + 1 = \boxed{3} \end{aligned}$$

c) For an arbitrary pmf $p_X(x)$ we have:

$$\begin{aligned} E\left[\frac{1}{f(x)}\right] &= \sum_{x\in\Omega_X} \frac{1}{f(x)} p_X(x) \\ &= \frac{1}{f(a)}p_X(a) + \frac{1}{f(b)}p_X(b) + \frac{1}{f(c)}p_X(c) \\ &= \frac{1}{10}p_X(a) + \frac{1}{5}p_X(b) + \frac{1}{10/7}p_X(c) \\ &= \frac{p_X(a)+2p_X(b)+7p_X(c)}{10} \\ &= \boxed{\frac{1+p_X(b)+6p_X(c)}{10}} \qquad \text{since } p_X(a)+p_X(b)+p_X(c) = 1 \end{aligned}$$

Problem 5: A biased four-sided die is rolled and the down face is a random variable N described by the following pmf:

$$p_N(n) = \begin{cases} n/10 & n = 1, 2, 3, 4 \\ 0 & \text{otherwise} \end{cases}$$

Given the random variable N a biased coin is flipped and the random variable X is 1 or zero according to whether the coin shows heads or tails. The conditional pmf is

$$p_{X|N}(x|n) = \left(\frac{n+1}{2n}\right)^x \left(1 - \frac{n+1}{2n}\right)^{1-x}$$

where $x \in \{0, 1\}$.

a) Find the expectation $E[N]$ and variance $V[N]$ of N.

b) Find the conditional pmf $p_{N|X}(n|x)$.

c) Find the conditional expectation $E[N|X = 1]$, i.e., the expectation with respect to the conditional pmf $p_{N|X}(n|1)$.

Solution: a)

$$
\begin{aligned}
E[N] &= \sum_{n\in\Omega_N} n \cdot p_N(n) && \text{by definition of expectation of a random variable}\\
&= 1\cdot p_N(1) + 2\cdot p_N(2) + 3\cdot p_N(3) + 4\cdot p_N(4) && \text{since } \Omega_N = \{1,2,3,4\}\\
&= 1\cdot 1/10 + 2\cdot 2/10 + 3\cdot 3/10 + 4\cdot 4/10 && \text{by pmf of } N\text{, i.e., } p_N\\
&= 1/10 + 4/10 + 9/10 + 16/10 && \text{arithmetic}\\
&= (1+4+9+16)/10 && \text{arithmetic}\\
&= 30/10 = \boxed{3}
\end{aligned}
$$

$$
\begin{aligned}
Var[N] &= E[N^2] - E[N]^2 && \text{by definition of Variance}
\end{aligned}
$$

Since we already know $E[N] = 3$, it follows, $E[N]^2 = 9$. Hence, we would only need to compute $E[N^2]$

$$
\begin{aligned}
E[N^2] &= \sum_{n\in\Omega_N} n^2 \cdot p_N(n) && \text{by definition of expectation of a random variable}\\
&= 1^2\cdot p_N(1) + 2^2\cdot p_N(2) + 3^2\cdot p_N(3) + 4^2\cdot p_N(4) && \text{since } \Omega_N = \{1,2,3,4\}\\
&= 1^2\cdot 1/10 + 2^2\cdot 2/10 + 3^2\cdot 3/10 + 4^2\cdot 4/10 && \text{by pmf of } N\text{, i.e., } p_N\\
&= 1/10 + 8/10 + 27/10 + 64/10 && \text{arithmetic}\\
&= (1+8+27+64)/10 && \text{arithmetic}\\
&= 100/10 = 10
\end{aligned}
$$

Therefore, we can compute the variance

$$
\begin{aligned}
Var[N] &= E[N^2] - E[N]^2 && \text{by definition of Variance}\\
&= 10 - 9 && \text{previously computed}\\
&= \boxed{1}
\end{aligned}
$$

b) By Bayes Rule, we have that

$$
p_{N|X}(n|x) = \frac{p_{X|N}(x|n)p_N(n)}{p_X(x)}
$$

To compute this pmf, we first need to compute the prior $p_X(x)$. We can easily find this as follow:

$$
\begin{aligned}
p_X(x=0) &= \sum_{n\in N} p_{X,N}(x=0,n) && \text{by marginalization}\\
&= \sum_{n\in N} p_{X|N}(x=0|n)p_N(n) && \text{by conditional probabilities}\\
&= p_{X|N}(x=0|n=1)p_N(n=1) + p_{X|N}(x=0|n=2)p_N(n=2) + \\
&\quad\; p_{X|N}(x=0|n=3)p_N(n=3) + p_{X|N}(x=0|n=4)p_N(n=4) && \text{expanding the sum}\\
&= 0\cdot\frac{1}{10} + \frac{1}{4}\cdot\frac{2}{10} + \frac{1}{3}\cdot\frac{3}{10} + \frac{3}{8}\cdot\frac{4}{10} && \text{using the given pmfs}\\
&= \frac{1}{10}\left[0 + \frac{2}{4} + \frac{3}{3} + \frac{12}{8}\right] && \text{arithmetic}\\
&= \frac{1}{10}[3] && \text{arithmetic}\\
&= \boxed{\frac{3}{10}}
\end{aligned}
$$

And

$$\begin{aligned}
p_X(x=1) &= \sum_{n \in N} p_{X,N}(x=1,n) && \text{by marginalization} \\
&= \sum_{n \in N} p_{X|N}(x=1|n)p_N(n) && \text{by conditional probabilities} \\
&= p_{X|N}(x=1|n=1)p_N(n=1) + p_{X|N}(x=1|n=2)p_N(n=2)+ \\
&\quad p_{X|N}(x=1|n=3)p_N(n=3) + p_{X|N}(x=1|n=4)p_N(n=4) && \text{expanding the sum} \\
&= 1 \cdot \frac{1}{10} + \frac{3}{4} \cdot \frac{2}{10} + \frac{2}{3} \cdot \frac{3}{10} + \frac{5}{8} \cdot \frac{4}{10} && \text{using the given pmfs} \\
&= \frac{1}{10}\left[1 + \frac{6}{4} + \frac{6}{3} + \frac{20}{8}\right] && \text{arithmetic} \\
&= \frac{1}{10}\left[\frac{14}{2}\right] && \text{arithmetic} \\
&= \boxed{\frac{7}{10}}
\end{aligned}$$

Note that since $X \in \{0,1\}$, we could have just computed $p_X(x=1) = 1 - p_X(x=0) = 1 - \frac{3}{10} = \frac{7}{10}$. Therefore, the pmf of X is given by:

$$p_X(x) = \begin{cases} \frac{3}{10} & \text{if } x = 0 \\ \frac{7}{10} & \text{otherwise, i.e., if } x = 1 \end{cases}$$

Now we have all the information we need to compute the pmf $p_{N|X}$. Let us compute one value at the time.

For $X = 0$:

$$\begin{aligned}
p_{N|X}(n=1|x=0) &= \frac{p_{X|N}(x=0|n=1) \cdot p_N(n=1)}{p_X(x=0)} && \text{by Bayes Rule} \\
&= \frac{0 \cdot \frac{1}{10}}{\frac{3}{10}} && \text{using the corresponding pmfs} \\
&= 0
\end{aligned}$$

$$\begin{aligned}
p_{N|X}(n=2|x=0) &= \frac{p_{X|N}(x=0|n=2) \cdot p_N(n=2)}{p_X(x=0)} && \text{by Bayes Rule} \\
&= \frac{\frac{1}{4} \cdot \frac{2}{10}}{\frac{3}{10}} && \text{using the corresponding pmfs} \\
&= \frac{1}{6}
\end{aligned}$$

$$\begin{aligned}
p_{N|X}(n=3|x=0) &= \frac{p_{X|N}(x=0|n=3) \cdot p_N(n=3)}{p_X(x=0)} && \text{by Bayes Rule} \\
&= \frac{\frac{1}{3} \cdot \frac{3}{10}}{\frac{3}{10}} && \text{using the corresponding pmfs} \\
&= \frac{1}{3}
\end{aligned}$$

$$\begin{aligned}
p_{N|X}(n=4|x=0) &= \frac{p_{X|N}(x=0|n=4)\cdot p_N(n=4)}{p_X(x=0)} && \text{by Bayes Rule} \\
&= \frac{\frac{3}{8}\cdot\frac{4}{10}}{\frac{3}{10}} && \text{using the corresponding pmfs} \\
&= \frac{1}{2}
\end{aligned}$$

Note that $0+\frac{1}{6}+\frac{1}{3}+\frac{1}{2}=\frac{1+2+3}{6}=\frac{6}{6}=1$

For $X=1$:

$$\begin{aligned}
p_{N|X}(n=1|x=1) &= \frac{p_{X|N}(x=1|n=1)\cdot p_N(n=1)}{p_X(x=1)} && \text{by Bayes Rule} \\
&= \frac{1\cdot\frac{1}{10}}{\frac{7}{10}} && \text{using the corresponding pmfs} \\
&= \frac{1}{7}
\end{aligned}$$

$$\begin{aligned}
p_{N|X}(n=2|x=1) &= \frac{p_{X|N}(x=1|n=2)\cdot p_N(n=2)}{p_X(x=1)} && \text{by Bayes Rule} \\
&= \frac{\frac{3}{4}\cdot\frac{2}{10}}{\frac{7}{10}} && \text{using the corresponding pmfs} \\
&= \frac{3}{14}
\end{aligned}$$

$$\begin{aligned}
p_{N|X}(n=3|x=1) &= \frac{p_{X|N}(x=1|n=3)\cdot p_N(n=3)}{p_X(x=1)} && \text{by Bayes Rule} \\
&= \frac{\frac{2}{3}\cdot\frac{3}{10}}{\frac{7}{10}} && \text{using the corresponding pmfs} \\
&= \frac{2}{7}
\end{aligned}$$

$$\begin{aligned}
p_{N|X}(n=4|x=1) &= \frac{p_{X|N}(x=1|n=4)\cdot p_N(n=4)}{p_X(x=1)} && \text{by Bayes Rule} \\
&= \frac{\frac{5}{8}\cdot\frac{4}{10}}{\frac{7}{10}} && \text{using the corresponding pmfs} \\
&= \frac{5}{14}
\end{aligned}$$

Note that $\frac{1}{7}+\frac{3}{14}+\frac{2}{7}+\frac{5}{14}=\frac{2+3+4+5}{14}=\frac{14}{14}=1$

Finally, let us summarize the pdf $p_{N|X}$ in the following table for easy display:

$n|x$	$n=1$	$n=2$	$n=3$	$n=4$
$x=0$	0	1/6	1/3	1/2
$x=1$	1/7	3/14	2/7	5/14

c)

$$
\begin{aligned}
E[N|X=1] &= \sum_{n\in N} n\cdot p_{N|X}(N=n|X=1) && \text{by conditional expectation}\\
&= 1\cdot p_{N|X}(N=1|X=1)+2\cdot p_{N|X}(N=2|X=1)\\
&\quad 3\cdot p_{N|X}(N=3|X=1)+4\cdot p_{N|X}(N=4|X=1) && \text{expanding the sum}\\
&= 1\cdot\frac{1}{7}+2\cdot\frac{3}{14}+3\cdot\frac{2}{7}+4\cdot\frac{5}{14} && \text{by the pmf } p_{N|X} \text{ calculated on b)}\\
&= \frac{2+6+12+20}{14} && \text{arithmetic}\\
&= \frac{20}{7}
\end{aligned}
$$

Therefore, $\boxed{E[N|X=1]=\frac{20}{7}}$

Problem 6: Let X, Y and Z be discrete random variables defined as functions on the same probability space $(\Omega, \mathcal{F}, P)$. Prove or disprove the following expression

$$P_{X|Y}(X=x|Y=y)=\sum_{z\in\Omega_Z} P_{X|YZ}(X=x|Y=y,Z=z)P_{Z|Y}(Z=z|Y=y)$$

Solution: This expression is true. The following is a proof:

$$
\begin{aligned}
\sum_{z\in\Omega_Z} P_{X|YZ}(X=x|Y=y,Z=z)P_{Z|Y}(Z=z|Y=y) &= \sum_{z\in\Omega_Z} \frac{P_{X|YZ}(X=x,Y=y,Z=z)}{P_{YZ}(Y=y,Z=z)}P_{Z|Y}(Z=z|Y=y)\\
&= \sum_{z\in\Omega_Z} \frac{P_{X|YZ}(X=x,Y=y,Z=z)}{P_{Z|Y}(Z=z|Y=Y)P_Y(Y=y)}P_{Z|Y}(Z=z|Y=y)\\
&= \sum_{z\in\Omega_Z} \frac{P_{X|YZ}(X=x,Y=y,Z=z)}{P_Y(Y=y)}\\
&= \sum_{z\in\Omega_Z} \frac{P_{X|YZ}(X=x,Z=z|Y=y)P_Y(Y=y)}{P_Y(Y=y)}\\
&= \sum_{z\in\Omega_Z} P_{X|YZ}(X=x,Z=z|Y=y)\\
&= P_{X|Y}(X=x|Y=y) \quad \text{by total probability}
\end{aligned}
$$

Note that in general the proof relies simply in the definition of conditional pmf, some arithmetic, and in total probability for the final equality where we note that z_i is a partition of Ω.

Problem 7: Consider the following game: A player is shown four closed doors and informed that the prize money is behind one of them (there is an equal chance that the money is behind each door). The player is asked to step in front of one of the doors, but then the following twist happens: Instead of revealing whether money is behind the door, the host chooses to open one of the remaining doors without the prize (the host picks the doors with equal probability). The player is then given an opportunity to step in front of some other door (he/she can choose to stay at the original door as well).

a) Define the probability space for this game. In particular, what is the sample space (hint: enumerate all outcomes of the game), what is the probability mass function, and consequently, what is the probability of winning the prize if the player stays at the original door.

b) Use random variables to solve this problem. What is the probability of winning the prize if the player stays at the original door and what is the probability of winning the prize if the player changes doors? Derive all expressions (no guesses) and clearly explain all random variables used in the derivation.

Solution: a) Let us define the sample space as follows:

$$\Omega = \{(P,C,O,S)|P,C,O,S \in \{1,2,3,4\} \text{ and } P \neq O \text{ and } S \neq O \text{ and } C \neq O\}$$

Here P stands for the door with the prize, C is the door chosen by the player, O is the door opened by the host and S is the door to which the player switched to (or possible not switch in case $C = S$).
For example: $(P,C,O,S) = (1,1,2,1) \in \Omega$ represents a game in which the prize was in door 1, the original guess of the player was door 1, the host opened door 2 but the player stayed in door 1. In this case the player won the prize. Another example, this time in which the player does not win would be $(P,C,O,S) = (3,2,4,1) \in \Omega$, representing a game in which the prize was in door 3, the player initial choice was door 2, the host opened door 4 and the player decided to move to door 1.

Note that $|\Omega| = 108$. We can see this with the following simple counting argument: choose a value for O. There are four possibilities. Once that value is fixed, we would have only 3 values for each of the other components P, C and S since $P \neq O$ and $S \neq O$ and $C \neq O$. Therefore, $|\Omega| = 3{\cdot}3{\cdot}4{\cdot}3 = 27{\cdot}4 = 108$

Since this is a discrete sample space, let us take $\mathcal{P}(\Omega) = \mathcal{F}$. To define the probability space $(\Omega, \mathcal{F}, P)$ it remains only to define P. As usual, we will instead define a probability mass function $p : \Omega \to [0,1]$ and define the probability of any event $A \in \mathcal{F}$ as $P(A) = \sum_{w \in A} p(w)$.

Define the pmf p, by assuming that the player stays at his original door. Then:

$$p((P,C,O,S)) = \begin{cases} 1/48 & (P,C,O,S) = (p,p,o,p), \text{ in words: the player initial choice is correct} \\ & \qquad\qquad \text{and he stays there, thereby winning the game} \\ 1/32 & (P,C,O,S) = (p,c,o,c), \text{ where } p \neq c. \text{ In words: the player initial choice is incorrect} \\ & \qquad\qquad \text{and he stays there, thereby loosing the game} \\ 0 & \text{otherwise} \end{cases}$$

It is easily seen that this is a valid pmf:

$$\sum_{w \in \Omega} p(w) = P(\text{initial choice correct OR initial choice incorrect}) = 4 \cdot \left(\frac{3}{48} + \frac{6}{32}\right) = 1$$

where 3/48 is, for a fixed door, the probability that the player initial choice is correct. But there are 4 possible doors, so me must multiply by 4 (assuming the initial price is at door $i \in \{1,2,3,4\}$ with probability 1/4. Also, 6/32 is the probability of the initial door incorrect where we again multiply by the number of doors, 4 in this case.

Finally, the probability of winning the prize if the player stays at the original door:

$$\begin{aligned} P(\text{winning the prize if the player stays at the original door}) &= P(\{w \in \Omega : w = (p,p,o,p) \text{where } p, o = 1,2,3,4\}) \\ &= P(\{(1,1,o,1) : o = 2,3,4\}) + \\ &\quad P(\{(2,2,o,2) : o = 1,3,4\}) + \\ &\quad P(\{(3,3,o,3) : o = 1,2,4\}) + \\ &\quad P(\{(4,4,o,4) : o = 1,2,3\}) \\ &= 4\left(\frac{3}{48}\right) \\ &= \frac{1}{4} \end{aligned}$$

Hence, the probability of winning if the player stays at the original door is just the same as at the beginning of the game 1/4.
A more interesting probability is that of winning if the player switches door. In the next part we verify that this probability is 3/8, better that 1/4, so it makes sense to change doors.

b) Let us define the following 3 random variables:

$$X = \begin{cases} 0 & \text{if price was at door originally picked by the player} \\ 1 & \text{otherwise} \end{cases}$$

With pmf: $P_X(X = 0) = 1/4$ and $P_X(X = 1) = 3/4$, since the prize is at a random location.

$$Y = \begin{cases} 0 & \text{if player switches to incorrect door} \\ 1 & \text{otherwise} \end{cases}$$

With pmf: $P_Y(Y = 0) = 1/2$ and $P_Y(Y = 1) = 1/2$, since at the time of switching there will be two choices, one with the prize and the other without prize.

Then we can compute the probabilities we want:

$$\begin{aligned} P(\text{player wins}|\text{player stays at original choice}) &= P_X(X=0) && \text{price at original choice} \\ &= 1/4 && \text{since prizes are randomly assigned} \end{aligned}$$

$$\begin{aligned} P(\text{player wins}|\text{player changes choice}) &= P_{XY}(X=1, Y=1) && \text{price not at original choice and switch correctly} \\ &= P_X(X=1) \cdot P_Y(Y=1) && \text{since these are independent} \\ &= 3/4 \cdot 1/2 && \text{by respective pmfs} \\ &= 3/8 && \text{arithmetic} \end{aligned}$$

We can see that switching is beneficial since $3/8 > 1/4$.
Note that X and Y are independent, because I am assuming that at the moment of switching doors the player will have 2 choices and he is going to pick one randomly regardless of where the price is since he does not know where the price is to begin with.

Problem 8: The time (in hours) necessary to find and fix an electrical problem in a certain institution is a random variable, say X, whose density is given by

$$p_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

If the cost of the breakdown of duration x is x^3, what is the expected cost of an electrical breakdown?

Solution: Let $C(x) = x^3$ be the cost function of the breakdown of duration x. Then, for the given random variable, the expected cost of an electrical breakdown is given by $E[C(x)]$. We can easily compute this:

$$\begin{aligned} E[C(x)] &= \int_{\Omega_X} C(x)p_X(x)dx && \text{by definition of expectation of a continuous random variable.} \\ &= \int_0^1 (x^3 \cdot 1)dx && \text{by definition of the random variable } X \text{ and } C(x) \\ &= \int_0^1 x^3 dx && \text{arithmetic} \\ &= \frac{x^4}{4}\Big|_0^1 && \text{solving the integral} \\ &= \boxed{\frac{1}{4}} && \text{where the expectation of } 1/4 \text{ is in units of money (whatever those are).} \end{aligned}$$

Problem 9: Let X be a continuos random variable with a cumulative distribution function $F_X(t)$. The median of a random variable is defined as a value of t for which

$$F_X(t) = \frac{1}{2}$$

Find the median of the random variables with the following density functions

a) $p_X(x) = e^{-x}, \qquad x > 0$
b) $p_X(x) = 1, \qquad 0 \leq x \leq 1.$

Solution: a) First, let us calculate the cdf of the random variable with given pdf

$$F_X(t) = \int_0^t p_X(x)dx = \int_0^t e^{-x}dx = -e^{-x}\Big|_0^t = -e^{-t} - (-e^0) = 1 - e^{-t} \qquad \text{for } t > 0$$

To solve for the median, set $F_X(t) = 1/2$ and solve for t as follow:

$$F_X(t) = 1/2 = 1 - e^{-t} \Longrightarrow 1/2 = e^{-t} \Longrightarrow ln(1/2) = -t \Longrightarrow -[ln(1) - ln(2)] = t \Longrightarrow \boxed{ln(2) = t}$$

b) Again, first, let us calculate the cdf of the random variable with given pdf

$$F_X(t) = \int_0^t p_X(x)dx = \int_0^t 1dx = x\Big|_0^t = t \qquad \text{for } 0 \leq t \leq 1$$

To solve for the median, set $F_X(t) = 1/2$ and solve for t as follow:

$$F_X(t) = \boxed{1/2 = t}$$

Extra P.: High dimensional spaces.

a) Show that in a high dimensional space, most of the volume of a cube is concentrated in corners, which themselves become very long "spikes." Hints: compute the ratio of the volume of a hypersphere of radius a to the volume of a hypercube of side $2a$ and also the ratio of the distance from the center of the hypercube to one of the corners divided by the perpendicular distance to one of the edges.

b) Show that for points which are uniformly distributed inside a sphere in d dimensions where d is large, almost all of the points are concentrated in a thin shell close to the surface. Hints: compute the fraction of the volume of the sphere which lies at values of the radius between $a - \epsilon$ and $0 < \epsilon < a$; Evaluate this fraction for $\epsilon = 0.01a$ and also for $\epsilon = 0.5a$ for $d \in \{2, 3, 10, 100\}$

Solution: a) By reference [7] (see end of this document), we know that the volume of a hypersphere in $n-$dimensions of radius a is given by:

$$B_a(n) = \frac{\pi^{n/2}a^n}{\Gamma\left(\frac{n}{2}+1\right)}, \qquad , \text{ where } \Gamma \text{ is the gamma function}$$

We also know that the volume of a hypercube in $n-$dimensions of side $2a$ is given by:

$$C_a(n) = (2a)^n$$

The ratio R of the volume of a hypersphere of radius a to the volume of a hypercube of side $2a$ is given by:

$$R(n) = \frac{\dfrac{\pi^{n/2}a^n}{\Gamma\left(\frac{n}{2}+1\right)}}{(2a)^n} = \frac{\pi^{n/2}a^n}{\Gamma\left(\frac{n}{2}+1\right)(2a)^n} = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)2^n}$$

The idea now is to explore what happens for high dimensional spaces, i.e., take the limit as n goes to infinity (in reality it would be the limit over integer values since dimensions are given by integers, but let us take limit over real values as a proxy for the result we want).

A simple calculation with wolfram alpha shows the result (see reference [3]), i.e.,

$$\lim_{n\to\infty} R(n) = \lim_{n\to\infty} \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)2^n} = 0$$

Therefore, for high values of n, the volume of the sphere is overwhelmed by the volume of the cube, and thus, most of the volume of a cube is concentrated in corners (since we think of the hypersphere as sitting inside the hypercube and touching its faces).

Now, let $D =$ distance from the center of the hypercube of side $2a$ to one of the corners. By repeatedly applying Pythagoras' theorem we see that

$$D = \sqrt{(2a)^2 + (2a)^2 + \cdots (2a)^2} = \sqrt{n(2a)^2} = 2a\sqrt{n}$$

Let $d =$ the perpendicular distance to one of the edges. This is given by:

$$d = 2a$$

Therefore, the ratio $R = D/d = \dfrac{2a\sqrt{n}}{2a} = \sqrt{n}$. Immediately wee see that:

$$\lim_{n\to\infty} R = \lim_{n\to\infty} \sqrt{n} = \infty$$

Implying that the distance from the center of the hypercube to one of the corners grows much faster than the perpendicular distance to one of the edges. This shows the result that in high dimensions the corners of the hypercube become very long "spikes."

b) Following the results in page 36 of the book (reference [8]), we first note that the volume of a sphere of radius a in d dimension is given by

$$V_d(a) = C_d a^d, \qquad \text{where } C_d \text{ is a constant that depends on } d \text{ alone and not on } a$$

The fraction of the volume of the sphere which lies at values of the radius between $a - \epsilon$ for $0 < \epsilon < a$ is

$$\frac{V_d(a) - V_d(a-\epsilon)}{V_d(a)} = \frac{C_d a^d - C_d(a-\epsilon)^d}{C_d a^d} = \frac{C_d(a^d - (a-\epsilon)^d)}{C_d a^d} = \frac{a^d - (a-\epsilon)^d}{a^d} = 1 - \frac{(a-\epsilon)^d}{a^d}$$

The following table summarizes the value of this fraction for for $\epsilon = 0.01a$ and also for $\epsilon = 0.5a$ for $d \in \{2, 3, 10, 100\}$:

$d \rightarrow$	2	3	10	100
$\epsilon = 0.01a$	0.0199	0.029701	0.0956179	0.633968
$\epsilon = 0.5a$	0.75	0.875	0.999023	≈ 1

From the table we can see that as d gets larger, the fraction of the points that lie in a thin shell close to the surface approaches 1, which means that almost all points are there! To complete this argument just note that the fraction tends to 1 as d approaches infinity since

$$\lim_{d\to\infty} 1 - \frac{(a-\epsilon)^d}{a^d} = \lim_{d\to\infty} 1 - \lim_{d\to\infty} \left(\frac{a-\epsilon}{a}\right)^d = 1 - 0 = 1$$

where the second limit is due to the fact that $0 < (a-\epsilon)/a < 1$ and hence, it approaches 0 as d approaches infinity. This proves the result: most of the points are in a thin shell, but only for spaces of high dimensions.

References

[1.] http://www.wolframalpha.com/input/?i=stirling+approximation+gamma+function

[2.] http://en.wikipedia.org/wiki/Stirling%27s_approximation#Stirling.27s_formula_for_the_gamma_function

[3.] http://www.wolframalpha.com/input/?i=limit+as+n+approaches+infinity
+%28pi%5E%28n%2F2%29%29%2F%28gamma%28%28n%2F2%29%2B1%292%5En%29

[4.] http://math.stackexchange.com/questions/608957/monty-hall-problem-extended

[5.] http://kobus.ca/seminars/ugrad/1-Introduction.pdf

[6.] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.156.41&rep=rep1&type=pdf

[7.] http://en.wikipedia.org/wiki/Volume_of_an_n-ball

[8.] http://kobus.ca/seminars/ugrad/1-Introduction.pdf