18.05 Problem Set 3, Spring 2014 Solutions

Problem 1. (10 pts.)

(a) We have $P(A) = P(B) = P(C) = 1/2$. Writing the outcome of die 1 first, we can easily list all outcomes in the following intersections.

$$A \cap B = \{(1,1), (1,3), (1,5), (3,1), (3,3), (3,5), (5,1), (5,3), (5,5)\}$$
$$A \cap C = \{(1,2), (1,4), (1,6), (3,2), (3,4), (3,6), (5,2), (5,4), (5,6)\}$$
$$B \cap C = \{(2,1), (4,1), (6,1), (2,3), (4,3), (6,3), (2,5), (4,5), (6,5)\}$$

By counting we see

$$P(A \cap B) = \frac{1}{4} = P(A)P(B).$$

Likewise,

$$P(A \cap C) = \frac{1}{4} = P(A)P(C) \quad \text{and} \quad P(B \cap C) = \frac{1}{4} = P(B)P(C).$$

So, we see that $A, B,$ and C are pairwise independent.

However, $A \cap B \cap C = \emptyset$, since if we roll an odd on die 1 and an odd on die 2, then the sum of the two will be even. So, in this case,

$$P(A \cap B \cap C) = 0 \neq P(A)P(B)P(C),$$

and we conclude that A, B and C are not mutually independent.

(b) By totaling the regions we get $P(A) = 0.225 + 0.05 + 0.1 + 0.125 = 0.5$. Likewise $P(B) = 0.5$ and $P(C) = 0.5$. Thus $P(A)P(B)P(C) = 0.5^3 = 0.125 = P(A \cap B \cap C)$. So, yes the product formula does hold.

Mutual independence requires pairwise independence as well as the multiplication formula for all three events. We see that

$$P(A \cap B) = 0.05 + 0.125 = 0.175, \text{ but } P(A)P(B) = 0.5^2 = 0.25.$$

Since $P(A)P(B) \neq P(A \cap B)$ the two events are not independent. However, $P(A)P(C) = 0.25$ and $P(A \cap C) = 0.225$, so A and C are not independent. Likewise $P(B)P(C) = 0.25$ and $= P(B \cap C) = 0.225$, so B and C are not independent.

Since the three events are not all pairwise independent they are not mutually independent.

(c) Let A be the event "the family has children of both sexes" and B be the event "there is at most one girl." In order for A to ever be true we assume that $n > 1$. Now, if we let X be the number of girls the we have

$$P(A) = P(1 \leq X \leq n-1) \qquad P(B) = P(X \leq 1) \qquad P(A \cap B) = P(X = 1)$$

Since $X \sim \text{binomial}(n, 1/2)$ we have

$$P(A) = 1 - P(X=0) - P(X=n) = 1 - \frac{2}{2^n}, \qquad P(B) = P(X=0) + P(X=1) = \frac{n+1}{2^n}.$$

Since we are told that A and B are independent, we have $P(A)P(B) = P(A \cap B)$, so

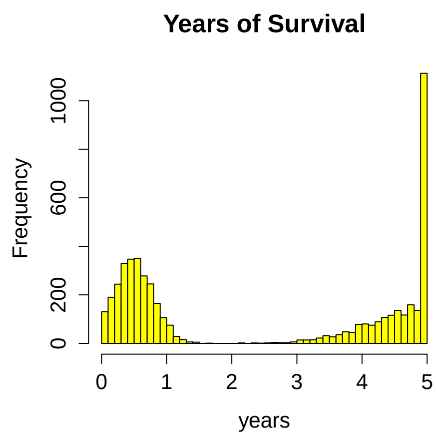
$$\begin{aligned} & \left(\tfrac{n+1}{2^n}\right)\left(1 - \tfrac{2}{2^n}\right) &&= \tfrac{n}{2^n} \\ \Leftrightarrow\quad & (n+1)\left(1 - \tfrac{2}{2^n}\right) &&= n \\ \Leftrightarrow\quad & n + 1 - \tfrac{n+1}{2^{n-1}} &&= n \\ \Leftrightarrow\quad & 2^{n-1} &&= n+1 \end{aligned}$$

Plugging small values of n into the above equation, we find that $n = 3$.

Problem 2. (10 pts.) For (a) and (b) the R-code is posted in ps3-sol.r

(a) mean = 2.554528, standard deviation = 2.07

(b)



(c) Looking at the distribution we see it is bimodal with a spike at 5 years. About half the patients die in the first year but about half live more than 2.5 years with over 20% still alive after 5 years. The spike is because everyone who survives to 5 years is lumped into that category. The average of 2.5 years is not that meaningful because there seem to be two categories of patients. This is reflected in the large standard deviation.

(d) The treatment appears to be effective for about half the patients. More research would be needed to understand what characteristics of the disease or patients predict the treatment will be effective.

Problem 3. (10 pts.)

(a) We compute $\text{Var}(X) = E(X^2) - E(X)$ etc. from the tables.

X	1	2	3	4
$p(x)$	1/4	1/4	1/4	1/4
X^2	1	4	9	16

Y	1	2	3	4	5	6
$p(y)$	1/6	1/6	1/6	1/6	1/6	1/6
Y^2	1	4	9	16	25	36

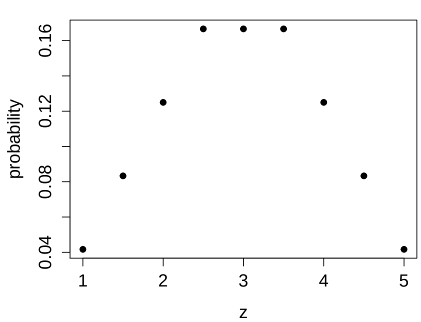
So, $E(X) = \frac{1}{4}(1+2+3+4) = \frac{5}{2}$, $E(X^2) = \frac{1}{4}(1+4+9+16) = \frac{15}{2}$. Thus, $\text{Var}(X) = 5/4 \;\Rightarrow\; \boxed{\sigma_X = \sqrt{5}/2}$.
Similarly, $E(Y) = \frac{7}{2}$, $E(X^2) = \frac{91}{6}$. So, $\text{Var}(Y) = \frac{35}{12}$.
Since X and Y are independent,

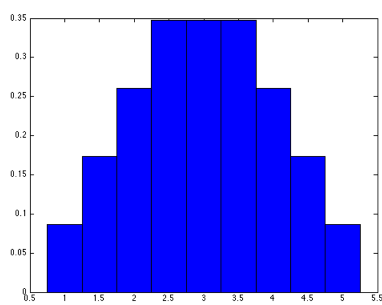
$$\text{Var}(Z) = \text{Var}\left(\frac{X+Y}{2}\right) = \frac{1}{4}\left(\text{Var}(X) + \text{Var}(Y)\right) = \frac{25}{24}.$$

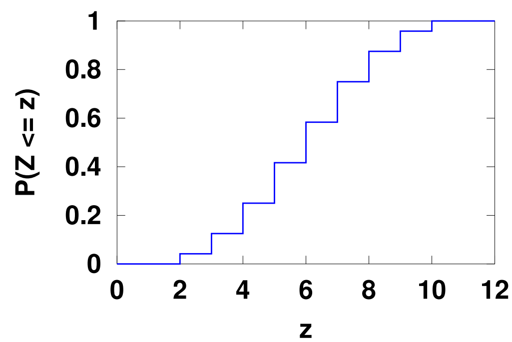
Thus,

$$\sigma_X = 1.118 \qquad \sigma_Y = 1.708 \qquad \sigma_Z = 1.021$$

(b) We graph the pmf of Z as point plot and then as a density histogram. The cdf is a staircase graph.







(c) We see that the only pairs of (X, Y) which satisfy $X > Y$ are $\{(2,1),(3,1),(3,2),(4,1),(4,2),(4,3)\}$. So $P(X > Y) = \frac{6}{24}$. Moreover, we have

$$P(X > Y|X = 2) = \frac{1}{6} \qquad P(X > Y|X = 3) = \frac{2}{6} \qquad P(X > Y|X = 4) = \frac{3}{6}$$

If W is our winnings for one game, we find

$$\begin{aligned} E(W) &= (-1)P(Y \geq X) + 2(2P(X > Y|X = 2)P(X = 2) + 3P(X > Y|X = 3)P(X = 3) + 4P(X > Y|X \\ &= -\frac{18}{24} + \frac{40}{24} \\ &= \frac{11}{12} \end{aligned}$$

Now if played the game 60 times, and received winnings $W_1, \ldots, W_{60}$, (with $E(W_i) = \frac{11}{12}$), our expected total gain is

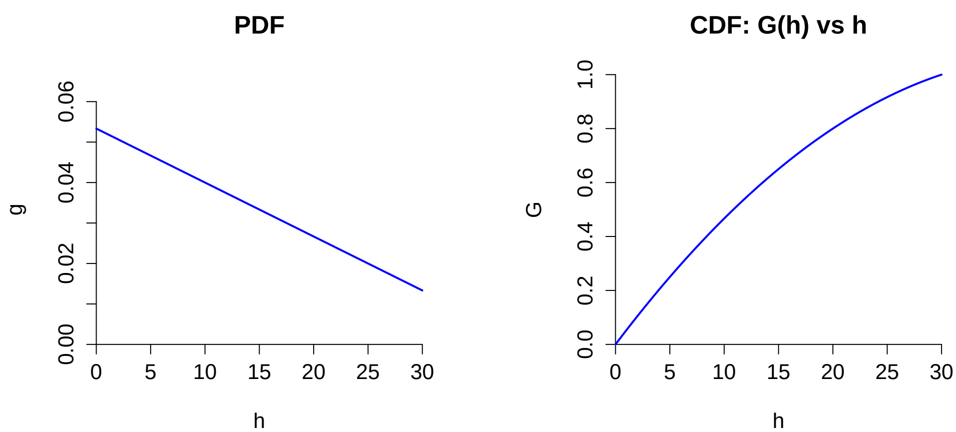
$$E(W_1 + \cdots + W_{60}) = E(W_1) + \cdots + E(W_{60}) = 55.$$

Problem 4. (10 pts.) (a) The number of raisins is

$$\int_0^{30} f(h)dh = \int_0^{30} (40-h)dh = 750$$

(b) The probability density is just the actual density divided by the total number of raisins. $g(h) = \frac{1}{750}(40-h)$.

(c) For $0 \le h \le 30$ we have $G(h) = \displaystyle\int_0^h g(x)\,dx = \frac{40h}{750} - \frac{h^2}{1500}$.



(d) Since the height is 30 we need to find $P(H \le 10)$.

$$P(H \le 10) = \int_0^{10} g(h)dh = \frac{1}{750}\int_0^{10}(40-h)dh = \frac{7}{15}.$$

The R code for these plots is posted in ps3-sol.r

Problem 5. (10 pts.)

(a) We know that $E(X) = aE(Z) + b = b$ and

$$\text{Var}(X) = \text{Var}(aZ+b) = a^2\text{Var}(Z) = a^2.$$

(b) Let x be any real number. We will first compute $F_X(x) = P(X \le x)$. Since $X = aZ + b$, we see that

$$F_X(x) = P(X \le x) = P(aZ + b \le x) = P\left(Z \le \frac{x-b}{a}\right) = \Phi\left(\frac{x-b}{a}\right).$$

So $F_X(x) = \Phi\left(\frac{x-b}{a}\right)$. Differentiating this with respect to x, we find

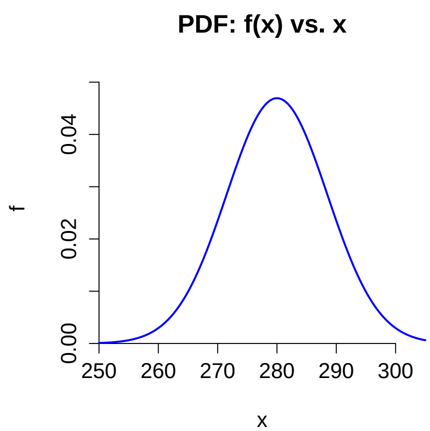
$$f_X(x) = \frac{1}{a}\Phi'\left(\frac{x-b}{a}\right) = \frac{1}{a}\phi\left(\frac{x-b}{a}\right) = \frac{1}{\sqrt{2\pi}\,a}\mathrm{e}^{-\frac{(x-b)^2}{2a^2}}$$

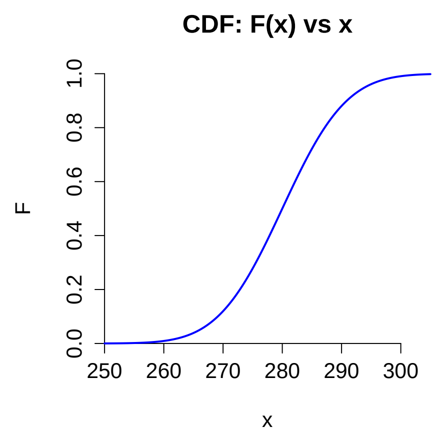
(c) From (b), we see that $f_X(x)$ is the pdf of $N(b, a^2)$ distribution

(d) From (b) and (c), we see that if Z is standard normal, then $\sigma Z + \mu$ follows a $N(\mu, \sigma^2)$ distribution. From (a), we know that $E(\sigma Z+\mu) = \mu$ and $\text{Var}(\sigma Z+\mu) = \sigma^2$.

Problem 6. (10 pts.)

(a) Suppose $Y \sim N(280, 8.5)$. The pdf, $f(y)$ and cdf $F(y)$ are plotted below.





(b) There is some ambiguity here depending on the exact time of day of the due date. On or before the day of the final means before midnight on the 18th. The due date is the 25th. We'll assume that means up to midnight, so the final is 7 days before the due date. We'll accept any number between 6 and 8

Let X be the number of days before or after May 25 that the baby is born. We want the probability $X \le -7$ We know $X \sim N(0, 8.5)$.

If Z is a standard normal random variable, we have

$$P(X \le -7) = \texttt{pnorm(-7,0,8.5)} = 0.205$$

(Or we could have computed $P(X \le -7) = P(Z \le -\frac{7}{8.5})$ `pnorm(-7/8.5,0,1)` $=$ 0.205.

(c) We want the probability that the baby is born between May 19 ($X = -6$) and May 31 ($X = 6$). We compute

$$P(-6 \le X \le 6) = P(-\frac{6}{8.5} \le X \le \frac{6}{8.5}) = 0.520$$

Again there is some ambiguity about the range. We'll accept any reasonable choice.

(d) We want to find x such that $P(X \ge x) > 0.95$. That is, we want $P(Z \ge \frac{x}{8.5}) \ge$ 0.95. Using R: `x = 8.5*qnorm(.05)`, we find $x \approx -14$ (May 11).

We could also have done the calculation with a standard normal table.

18.05 Introduction to Probability and Statistics
Spring 2014