Class 26: review for final exam –solutions, 18.05, Spring 2014

Problem 1. **(a)** Four ways to fill each slot: $\boxed{4^5}$.

(b) Four ways to fill the first slot and 3 ways to fill each subsequent slot: $\boxed{4 \cdot 3^4}$.

(c) Build the sequences as follows:
Step 1: Choose which of the 5 slots gets the A: 5 ways to place the one A.
Step 2: 3^4 ways to fill the remain 4 slots. By the rule of product there are $\boxed{5 \cdot 3^4}$ such sequences.

Problem 2. **(a)** $\binom{52}{5}$.

(b) Number of ways to get a full-house: $\binom{4}{2}\binom{13}{1}\binom{4}{3}\binom{12}{1}$

(c) $\dfrac{\binom{4}{2}\binom{13}{1}\binom{4}{3}\binom{12}{1}}{\binom{52}{5}}$

Problem 3. **(a)** There are several ways to think about this. Here is one.

The 11 letters are p, r, o, b,b, a, i,i, l, t, y. We use the following steps to create a sequence of these letters.
Step 1: Choose a position for the letter p: 11 ways to do this.
Step 2: Choose a position for the letter r: 10 ways to do this.
Step 3: Choose a position for the letter o: 9 ways to do this.
Step 4: Choose two positions for the two b's: 8 choose 2 ways to do this.
Step 5: Choose a position for the letter a: 6 ways to do this.
Step 6: Choose two positions for the two i's: 5 choose 2 ways to do this.
Step 7: Choose a position for the letter l: 3 ways to do this.
Step 8: Choose a position for the letter t: 2 ways to do this.
Step 9: Choose a position for the letter y: 1 ways to do this.

Multiply these all together we get:

$$11 \cdot 10 \cdot 9 \cdot \binom{8}{2} \cdot 6 \cdot \binom{5}{2} \cdot 3 \cdot 2 \cdot 1 \;=\; \frac{11!}{2! \cdot 2!}$$

(b) Here are two ways to do this problem.

Method 1. Since every arrangement has equal probability of being chosen we simply have to count the number that start with the letter 'b'. After putting a 'b' in position 1 there are 10 letters: p, r, o, b, a, i,i, l, t, y, to place in the last 10 positions. We count this in the same manner as part (a). That is
Choose the position for p: 10 ways.
Choose the positions for r,o,b,a,: $9 \cdot 8 \cdot 7 \cdot 6$ ways.
Choose two positions for the two i's: 5 choose 2 ways.
Choose the position for l: 3 ways.

Choose the position for t: 2 ways.
Choose the position for y: 1 ways.

Multiplying this together we get $10\cdot 9\cdot 8\cdot 7\cdot 6\cdot \binom{5}{2}\cdot 3\cdot 2\cdot 1 = \dfrac{10!}{2!}$ arrangements start with the letter b. Therefore the probability a random arrangement starts with b is $\boxed{\dfrac{10!/2!}{11!/2!\cdot 2!} = \dfrac{2}{11}}$

Method 2. Suppose we build the arrangement by picking a letter for the first position, then the second position etc. Since there are 11 letters, two of which are b's we have a 2/11 chance of picking a b for the first letter.

Problem 4. We are given $P(E \cup F) = 2/3$.
$E^c \cap F^c = (E \cup F)^c \;\Rightarrow\; P(E^c \cap F^c) = 1 - P(E \cup F) = \boxed{1/3.}$

Problem 5.

D is the disjoint union of $D \cap C$ and $D \cap C^c$.
So, $P(D \cap C) + P(D \cap C^c) = P(D)$
$\Rightarrow\; P(D \cap C) = P(D) - P(D \cap C^c) = .4 - .2 = \boxed{.2.}$

Problem 6. **(a)** Slots 1, 3, 5, 7 are filled by T_1, T_3, T_5, T_7 in any order: 4! ways.
Slots 2, 4, 6, 8 are filled by T_2, T_4, T_6, T_8 in any order: 4! ways.
answer: $4!\cdot 4! = 576$.

(b) There are 8! ways to fill the 8 slots in any way.
Since each outcome is equally likely the probabilitiy is
$\dfrac{4!\cdot 4!}{8!} = \dfrac{576}{40320} = 0.143 = 1.43\%$.

Problem 7. Let H_i be the event that the i^{th} hand has one king. We have the conditional probabilities

$$P(H_1) = \frac{\binom{4}{1}\binom{48}{12}}{\binom{52}{13}}; \quad P(H_2|H_1) = \frac{\binom{3}{1}\binom{36}{12}}{\binom{39}{13}}; \quad P(H_3|H_1 \cap H_2) = \frac{\binom{2}{1}\binom{24}{12}}{\binom{26}{13}}$$

$P(H_4|H_1 \cap H_2 \cap H_3) = 1$

$$\begin{aligned} P(H_1 \cap H_2 \cap H_3 \cap H_4) &= P(H_4|H_1 \cap H_2 \cap H_3)\, P(H_3|H_1 \cap H_2)\, P(H_2|H_1)\, P(H_1) \\ &= \frac{\binom{2}{1}\binom{24}{12}\binom{3}{1}\binom{36}{12}\binom{4}{1}\binom{48}{12}}{\binom{26}{13}\binom{39}{13}\binom{52}{13}}. \end{aligned}$$

Problem 8.

(a) Sample space $= \Omega = \{(1,1), (1,2), (1,3), \ldots, (6,6)\} = \{(i,j) \,|\, i,j = 1, 2, 3, 4, 5, 6\}$.
(Each outcome is equally likely, with probability 1/36.)
$A = \{(1,4), (2,3), (3,2), (4,1)\}$,
$B = \{(4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (1,4), (2,4), (3,4), (5,4), (6,4)\}$
$P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{2/36}{11/36} = \dfrac{2}{11}..$

(b) $P(A) = 4/36 \neq P(A|B)$, so they are not independent.

Problem 9. Let C be the event the contestant gets the question correct and G the event the contestant guessed.

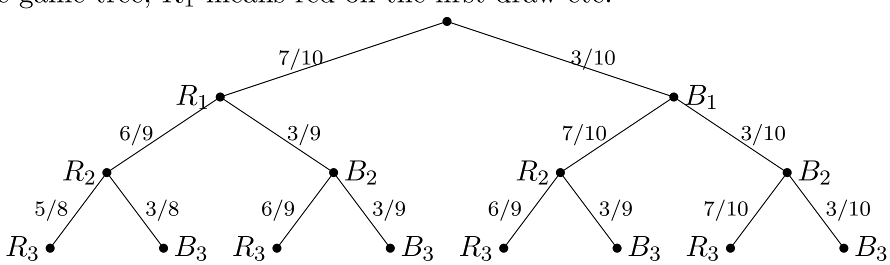
The question asks for $P(G|C)$.

We'll compute this using Bayes' rule: $P(G|C) = \dfrac{P(C|G)\,P(G)}{P(C)}$.

We're given: $P(C|G) = 0.25$, $P(K) = 0.7$.
Law of total prob.:
$P(C) = P(C|G)\,P(G) + P(C|G^c)\,P(G^c) = 0.25 \cdot 0.3 + 1.0 \cdot 0.7 = 0.775$.
Therefore $P(G|C) = \dfrac{0.075}{0.775} = 0.097 = 9.7\%$.

Problem 10.

Here is the game tree, R_1 means red on the first draw etc.



Summing the probability to all the B_3 nodes we get
$P(B_3) = \dfrac{7}{10} \cdot \dfrac{6}{9} \cdot \dfrac{3}{8} + \dfrac{7}{10} \cdot \dfrac{3}{9} \cdot \dfrac{3}{9} + \dfrac{3}{10} \cdot \dfrac{7}{10} \cdot \dfrac{3}{9} + \dfrac{3}{10} \cdot \dfrac{3}{10} \cdot \dfrac{3}{10} = .350.$

Problem 11. We have $P(A \cup B) = 1 - 0.42 = 0.58$ and we know

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

Thus,

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.4 + 0.3 - 0.58 = 0.12 = (0.4)(0.3) = P(A)P(B)$$

So A and B are independent.

Problem 12. We have

$$P(A \cap B \cap C) = 0.06 \qquad P(A \cap B) = 0.12$$
$$P(A \cap C) = 0.15 \qquad P(B \cap C) = 0.2$$

Since $P(A \cap B) = P(A \cap B \cap C) + P(A \cap B \cap C^c)$, we find $P(A \cap B \cap C^c) = 0.06$. Similarly

$$
\begin{aligned}
P(A \cap B \cap C^c) &= 0.06 \\
P(A \cap B^c \cap C) &= 0.09 \\
P(A^c \cap B \cap C) &= 0.14
\end{aligned}
$$

Problem 13. To show A and B are not independent we need to show $P(A \cap B) \neq P(A) \cdot P(B)$.

(a) No, they cannot be independent: $A \cap B = \emptyset \;\Rightarrow\; P(A \cap B) = 0 \neq P(A) \cdot P(B)$.

(b) No, they cannot be independent: same reason as in part (a).

Problem 14.

X	-2	-1	0	1	2
p(X)	1/15	2/15	3/15	4/15	5/15

We compute

$$E[X] = -2 \cdot \frac{1}{15} + -1 \cdot \frac{2}{15} + 0 \cdot \frac{3]15 + 1 \cdot \frac{[}{4}}{15} + 2 \cdot \frac{5}{15} = \frac{2}{3}.$$

Thus

$$\mathrm{Var}(X) = E((X - \frac{2}{3})^2) = \frac{14}{9}.$$

Problem 15. We first compute

$$
\begin{aligned}
E[X] &= \int_0^1 x \cdot 2x dx = \frac{2}{3} \\
E[X^2] &= \int_0^1 x^2 \cdot 2x dx = \frac{1}{2} \\
E[X^4] &= \int_0^1 x^4 \cdot 2x dx = \frac{1}{3}.
\end{aligned}
$$

Thus,

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

and

$$\mathrm{Var}(X^2) = E[X^4] = \left(E[X^2]\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Problem 16. **(a)** We have

X values:	-1	0	1
prob:	1/3	1/6	1/2
X^2	1	0	1

So, $\boxed{E(X) = -1/3 + 1/2 = 1/6.}$

(b) $\begin{array}{lcc} Y \text{ values:} & 0 & 1 \\ \text{prob:} & 1/6 & 5/6 \end{array} \Rightarrow \boxed{E(Y) = 5/6.}$

(c) Using the table in part (a) $E(X^2) = 1 \cdot (1/3) + 0 \cdot (1/6) + 1 \cdot (1/2) = 5/6$ (same as part (b)).

(d) $\boxed{\text{Var}(X) = E(X^2) - E(X)^2 = 5/6 - 1/36 = 29/36.}$

Problem 17. **answer:**

Make a table

X:	0	1
prob:	(1-p)	p
X^2	0	1.

From the table, $\boxed{E(X) = 0 \cdot (1-p) + 1 \cdot p = p.}$

Since X and X^2 have the same table $E(X^2) = E(X) = p$.

Therefore, $\boxed{\text{Var}(X) = p - p^2 = p(1-p).}$

Problem 18. Let X be the number of people who get their own hat.

Following the hint: let X_j represent whether person j gets their own hat. That is, $X_j = 1$ if person j gets their hat and 0 if not.

We have, $X = \sum_{j=1}^{100} X_j$, so $E(X) = \sum_{j=1}^{100} E(X_j)$.

Since person j is equally likely to get any hat, we have $P(X_j = 1) = 1/100$. Thus, $X_j \sim$ Bernoulli$(1/100) \Rightarrow E(X_j) = 1/100 \Rightarrow \boxed{E(X) = 1.}$

Problem 19. For $y = 0, 2, 4, \ldots, 2n$,

$$P(Y = y) = P(X = \frac{y}{2}) = \binom{n}{y/2} \frac{1}{2}^n.$$

Problem 20. We have $f_X(x) = 1$ for $0 \le x \le 1$. The cdf of X is

$$F_X(x) = \int_0^x f_X(t)dt = \int_0^x 1dt = x.$$

Now for $5 \le y \le 7$, we have

$$F_Y(y) = P(Y \le y) = P(2X + 5 \le y) = P(X \le \frac{y-5}{2}) = F_X(\frac{y-5}{2}) = \frac{y-5}{2}.$$

Differentiating $P(Y \le y)$ with respect to y, we get the probability density function of Y, for $5 \le y \le 7$,

$$f_Y(y) = \frac{1}{2}.$$

Problem 21. We have cdf of X,

$$F_X(x) = \int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}.$$

Now for $y \geq 0$, we have

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = 1 - e^{-\lambda\sqrt{y}}.$$

Differentiating $F_Y(y)$ with respect to y, we have

$$f_Y(y) = \frac{\lambda}{2} y^{-\frac{1}{2}} e^{-\lambda\sqrt{y}}.$$

Problem 22. **(a)** We first make the probability tables

X	0	2	3
prob.	0.3	0.1	0.6
Y	3	3	12

$\Rightarrow\ E(X) = 0 \cdot 0.3 + 2 \cdot 0.1 + 3 \cdot 0.6 = 2$

(b) $E(X^2) = 0 \cdot 0.3 + 4 \cdot 0.1 + 9 \cdot 0.6 = 5.8 \ \Rightarrow\ \mathrm{Var}(X) = E(X^2) - E(X)^2 = 5.8 - 4 = 1.8.$

(c) $E(Y) = 3 \cdot 0.3 + 3 \cdot 0.1 + 12 \cdot 6 = 8.4.$

L(d) $F_Y(7) = P(Y \leq 7) = 0.4.$

Problem 23.

(a) There are a number of ways to present this.

$X \sim 3\,\mathrm{binomial}(25, 1/6)$, so

$$P(X = 3k) = \binom{25}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{25-k}, \quad \text{for } k = 0,\ 1,\ 2,\ \ldots,\ 25.$$

(b) $X \sim 3\,\mathrm{binomial}(25, 1/6)$.

Recall that the mean and variance of binomial(n, p) are np and $np(1-p)$. So,

$$E(X) = 3np = 3 \cdot 25 \cdot \frac{1}{6} = 75/6, \text{ and } \mathrm{Var}(X) = 9np(1-p) = 9 \cdot 25(1/6)(5/6) = 125/4.$$

(c) $E(X+Y) = E(X) + E(Y) = 150/6 = 25.$, $E(2X) = 2E(X) = 150/6 = 25.$

$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 250/4$. $\mathrm{Var}(2X) = 4\mathrm{Var}(X) = 500/4.$

The means of $X+Y$ and $2X$ are the same, but $\mathrm{Var}(2X) > \mathrm{Var}(X+Y)$.

This makes sense because in $X+Y$ sometimes X and Y will be on opposite sides from the mean so distances to the mean will tend to cancel, However in $2X$ the distance to the mean is always doubled.

Problem 24. First we find the value of a:

$$\int_0^1 f(x)\,dx = 1 = \int_0^1 x + ax^2\,dx = \frac{1}{2} + \frac{a}{3} \;\Rightarrow\; a = 3/2.$$

The CDF is $F_X(x) = P(X \le x)$. We break this into cases:
(i) $b < 0 \;\Rightarrow\; F_X(b) = 0$.
(ii) $0 \le b \le 1 \;\Rightarrow\; F_X(b) = \displaystyle\int_0^b x + \frac{3}{2}x^2\,dx = \frac{b^2}{2} + \frac{b^3}{2}$.
(iii) $1 < x \;\Rightarrow\; F_X(b) = 1$.
Using F_X we get

$$P(.5 < X < 1) = F_X(1) - F_X(.5) = 1 - \left(\frac{.5^2 + .5^3}{2}\right) = \frac{13}{16}.$$

Problem 25.
(i) yes, discrete, (ii) no, (iii) no, (iv) no, (v) yes, continuous
(vi) no (vii) yes, continuous, (viii) yes, continuous.

Problem 26.
(a) We compute

$$P(X \ge 5) = 1 - P(X < 5) = 1 - \int_0^5 \lambda e^{-\lambda x} dx = 1 - (1 - e^{-5\lambda}) = e^{-5\lambda}.$$

(b) We want $P(X \ge 15 | X \ge 10)$. First observe that $P(X \ge 15, X \ge 10) = P(X \ge 15)$. From similar computations in (a), we know

$$P(X \ge 15) = e^{-15\lambda} \qquad\qquad P(X \ge 10) = e^{-10\lambda}.$$

From the definition of conditional probability,

$$P(X \ge 15 | X \ge 10) = \frac{P(X \ge 15, X \ge 10)}{P(X \ge 10)} = \frac{P(X \ge 15)}{P(X \ge 10)} = e^{-5\lambda}$$

Note: This is an illustration of the memorylessness property of the exponential distribution.

Problem 27. **(a)** We have

$$F_X(x) = P(X \le x) = P(3Z + 1 \le x) = P(Z \le \frac{x-1}{3}) = \Phi(\frac{x-1}{3}).$$

(b) Differentiating with respect to x, we have

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{dx}} F_X(x) = \frac{1}{3}\phi(\frac{x-1}{3}).$$

Since $\phi(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}}$, we conclude

$$f_X(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{(x-1)^2}{2\cdot 3^2}},$$

which is the probability density function of the $N(1, 9)$ distribution. **Note:** The arguments in (a) and (b) give a proof that $3Z+1$ is a normal random variable with mean 1 and variance 9. See Problem Set 3, Question 5.

(c) We have

$$P(-1 \leq X \leq 1) = P(-\frac{2}{3} \leq Z \leq 0) = \Phi(0) - \Phi(-\frac{2}{3}) \approx 0.2475$$

(d) Since $E(X) = 1$, $\text{Var}(X) = 9$, we want $P(-2 \leq X \leq 4)$. We have

$$P(-2 \leq X \leq 4) = P(-3 \leq 3Z \leq 3) = P(-1 \leq Z \leq 1) \approx 0.68.$$

Problem 28. **(a)** Note, Y follows what is called a *log-normal distribution.*
$F_Y(a) = P(Y \leq a) = P(e^Z \leq a) = P(Z \leq \ln(a)) = \boxed{\Phi(\ln(a)).}$
Differentiating using the chain rule:

$$f_y(a) = \frac{d}{da} F_Y(a) = \frac{d}{da} \Phi(\ln(a)) = \frac{1}{a} \phi(\ln(a)) = \boxed{\frac{1}{\sqrt{2\pi}\, a} e^{-(\ln(a))^2/2}.}$$

(b) (i) We want to find $q_{.33}$ such that $P(Z \leq q_{.33}) = .33$. That is, we want

$$\Phi(q_{.33}) = .33 \;\Leftrightarrow\; \boxed{q_{.33} = \Phi^{-1}(.33)}.$$

(ii) We want $q_{.9}$ such that

$$F_Y(q_{.9}) = .9 \;\Leftrightarrow\; \Phi(\ln(q_{.9})) = .9 \;\Leftrightarrow\; \boxed{q_{.9} = e^{\Phi^{-1}(.9)}}.$$

(iii) As in (ii) $q_{.5} = e^{\Phi^{-1}(.5)} = e^0 = \boxed{1}$.

Problem 29. **(a)** <u>**answer:**</u> $\text{Var}(X_j) = 1 = E(X_j^2) - E(X_j)^2 = E(X_j^2)$. QED

(b) $E(X_j^4) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\infty} x^4 e^{-x^2/2}\, dx.$

(Extra credit) By parts: let $u = x^3$, $v' = xe^{-x^2/2} \;\Rightarrow\; u' = 3x^2$, $v = -e^{-x^2/2}$

$E(X_j^4) = \dfrac{1}{\sqrt{2\pi}} \left[x^3 e^{-x^2/2} \Big|_{infty}^{\infty} + \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\infty} 3x^2 e^{-x^2/2}\, dx \right]$

The first term is 0 and the second term is the formula for $3E(X_j^2) = 3$ (by part (a)). Thus, $E(X_j^4) = 3$.

(c) <u>**answer:**</u> $\text{Var}(X_j^2) = E(X_j^4) - E(X_j^2)^2 = 3 - 1 = 2$. QED

(d) $E(Y_{100}) = E(100X_j^2) = 100$. $\text{Var}(Y_{100}) = 100\text{Var}(X_j) = 200$.
The CLT says Y_{100} is approximately normal. Standardizing gives
$P(Y_{100} > 110) = P\left(\frac{Y_{100}-100}{\sqrt{200}}) > \frac{10}{\sqrt{200}}\right) \approx P(Z > 1/\sqrt{2}) = \boxed{.24}$.
This last value was computed using `1 - pnorm(1/sqrt(2),0,1)`.

Problem 30.

(a) We did this in class. Let $\phi(z)$ and $\Phi(z)$ be the PDF and CDF of Z.
$F_Y(y) = P(Y \le y) = P(aZ + b \le y) = P(Z \le (y-b)/a) = \Phi((y-b)/a)$.
Differentiating:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}\Phi((y-b)/a) = \frac{1}{a}\phi((y-b)/a) = \frac{1}{\sqrt{2\pi}\,a}\,\mathrm{e}^{-(y-b)^2/2a^2}.$$

Since this is the density for $\mathrm{N}(b, a^2)$ we have shown $Y \sim \mathrm{N}(b, a^2)$.

(b) By part (a), $Y \sim \mathrm{N}(\mu, \sigma^2) \Rightarrow Y = \sigma Z + \mu$.
But, this implies $(Y - \mu)/\sigma = Z \sim \mathrm{N}(0, 1)$. QED

Problem 31.

(a) $E(W) = 3E(X) - 2E(Y) + 1 = 6 - 10 + 1 = \boxed{-3}$
$\text{Var}(W) = 9\text{Var}(X) + 4\text{Var}(Y) = 45 + 36 = \boxed{81}$

(b) Since the sum of independent normal is normal part (a) shows: $W \sim N(-3, 81)$. Let $Z \sim N(0, 1)$. We *standardize* W: $P(W \le 6) = P\left(\frac{W+3}{9} \le \frac{9}{9}\right) = P(Z \le 1) \approx \boxed{.84.}$

Problem 32.

Method 1

$U(a, b)$ has density $f(x) = \frac{1}{b-a}$ on $[a, b]$. So,

$$E(X) = \int_a^b x f(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b^2-a^2}{2(b-a)} = \boxed{\frac{a+b}{2}.}$$

$$E(X^2) = \int_a^b x^2 f(x)\,dx = \frac{1}{b-a}\int_a^b x^2\,dx = \left.\frac{x^3}{3(b-a)}\right|_a^b = \frac{b^3-a^3}{3(b-a)}.$$

Finding $\text{Var}(X)$ now requires a little algebra,

$$\begin{aligned}\text{Var}(X) = E(X^2) - E(X)^2 &= \frac{b^3-a^3}{3(b-a)} - \frac{(b+a)^2}{4} \\ &= \frac{4(b^3-a^3) - 3(b-a)(b+a)^2}{12(b-a)} = \frac{b^3 - 3ab^2 + 3a^2b - a^3}{12(b-a)} = \frac{(b-a)^3}{12(b-a)} = \boxed{\frac{(b-a)^2}{12}.}\end{aligned}$$

Method 2

There is an easier way to find $E(X)$ and $\text{Var}(X)$.

Let $U \sim \text{U}(0,1)$. Then the calculations above show $E(U) = 1/2$ and $(E(U^2) = 1/3$ $\Rightarrow$ $\text{Var}(U) = 1/3 - 1/4 = 1/12$.
Now, we know $X = (b-a)U + a$, so $E(X) = (b-a)E(U) + a = (b-a)/2 + a = (b+a)/2$ and $\text{Var}(X) = (b-a)^2\text{Var}(U) = (b-a)^2/12$.

Problem 33.

(a) $S_n \sim \text{Binomial}(n, p)$, since it is the number of successes in n independent Bernoulli trials.

(b) $T_m \sim \text{Binomial}(m, p)$, since it is the number of successes in m independent Bernoulli trials.

(c) $S_n + T_m \sim \text{Binomial}(n+m, p)$, since it is the number of successes in $n+m$ independent Bernoulli trials.

(d) Yes, S_n and T_m are independent since trials 1 to n are independent of trials $n+1$ to $n+m$.

Problem 34. Compute the median for the exponential distribution with parameter λ. The density for this distribution is $f(x) = \lambda\,\mathrm{e}^{-\lambda x}$. We know (or can compute) that the distribution function is $F(a) = 1 - \mathrm{e}^{-\lambda a}$. The median is the value of a such that $F(a) = .5$. Thus, $1 - \mathrm{e}^{-\lambda a} = 0.5 \;\Rightarrow\; 0.5 = \mathrm{e}^{-\lambda a} \;\Rightarrow\; \log(0.5) = -\lambda a \;\Rightarrow\; \boxed{a = \log(2)/\lambda.}$

Problem 35. Let X = the number of heads on the first 2 flips and Y the number in the last 2. Considering all 8 possibe tosses: HHH, HHT etc we get the following joint pmf for X and Y

Y/X	0	1	2	
0	1/8	1/8	0	1/4
1	1/8	1/4	1/8	1/2
2	0	1/8	1/8	1/4
	1/4	1/2	1/4	1

Using the table we find

$$E(XY) = \frac{1}{4} + 2\frac{1}{8} + 2\frac{1}{8} + 4\frac{1}{8} = \frac{5}{4}.$$

We know $E(X) = 1 = E(Y)$ so

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{5}{4} - 1 = \frac{1}{4}.$$

Since X is the sum of 2 independent Bernoulli(.5) we have $\sigma_X = \sqrt{2/4}$

$$\text{Cor}(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{1/4}{(2)/4} = \frac{1}{2}.$$

Problem 36. As usual let X_i = the number of heads on the i^{th} flip, i.e. 0 or 1.

Let $X = X_1 + X_2 + X_3$ the sum of the first 3 flips and $Y = X_3 + X_4 + X_5$ the sum of the last 3. Using the algebraic properties of covariance we have

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathrm{Cov}(X_1 + X_2 + X_3, X_3 + X_4 + X_5) \\
&= \mathrm{Cov}(X_1, X_3) + \mathrm{Cov}(X_1, X_4) + \mathrm{Cov}(X_1, X_5) \\
&+ \mathrm{Cov}(X_2, X_3) + \mathrm{Cov}(X_2, X_4) + \mathrm{Cov}(X_2, X_5) \\
&+ \mathrm{Cov}(X_3, X_3) + \mathrm{Cov}(X_3, X_4) + \mathrm{Cov}(X_3, X_5)
\end{aligned}$$

Because the X_i are independent the only non-zero term in the above sum is $\mathrm{Cov}(X_3 X_3) = \mathrm{Var}(X_3) = \dfrac{1}{4}$
Therefore, $\mathrm{Cov}(X, Y) = \frac{1}{4}$.

We get the correlation by dividing by the standard deviations. Since X is the sum of 3 independent Bernoulli(.5) we have $\sigma_X = \sqrt{3/4}$

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{1/4}{(3)/4} = \frac{1}{3}.$$

Problem 37. **(a)** X and Y are independent, so the table is computed from the product of the known marginal probabilities. Since they are independent, $\mathrm{Cov}(X, Y) = 0$.

$Y\backslash X$	0	1	P_Y
0	1/8	1/8	1/4
1	1/4	1/4	1/2
2	1/8	1/8	1/4
P_X	1/2	1/2	1

(b) The sample space is $\Omega = \{$HHH, HHT, HTH, HTT, THH, THT, TTH, TTT$\}$.
$P(X = 0, Z = 0) = P(\{TTH, TTT\}) = 1/4$.
$P(X = 0, Z = 1) = P(\{THH, THT\}) = 1/4$.
$P(X = 0, Z = 2) = 0$.
$P(X = 1, Z = 0) = 0$.
$P(X = 1, Z = 1) = P(\{HTH, HTT\}) = 1/4$.
$P(X = 1, Z = 2) = P(\{HHH, HHT\}) = 1/4$.

$Z\backslash X$	0	1	P_Z
0	1/4	0	1/4
1	1/4	1/4	1/2
2	0	1/4	1/4
P_X	1/2	1/2	1

$\mathrm{Cov}(X, Z) = E(XZ) - E(X)E(Z)$.
$E(X) = 1/2, \quad E(Z) = 1, \; E(XZ) = \sum x_i y_j p(x_i, y_j) = 3/4$.
$\Rightarrow \; \mathrm{Cov}(X, Z) = 3/4 - 1/2 = \boxed{1/4.}$

Problem 38. **(a)**

$Y \backslash X$	-2	-1	0	1	2	
0	0	0	1/5	0	0	1/5
1	0	1/5	0	1/5	0	2/5
4	1/5	0	0	0	1/5	2/5
	1/5	1/5	1/5	1/5	1/5	1

Each column has only one nonzero value. For example, when $X = -2$ then $Y = 4$, so in the $X = -2$ column, only $P(X = -2, Y = 4)$ is not 0.

(b) Using the marginal distributions: $E(X) = \frac{1}{5}(-2 - 1 + 0 + 1 + 2) = 0$.

$E(Y) = 0 \cdot \dfrac{1}{5} + 1 \cdot \dfrac{2}{5} + 4 \cdot \dfrac{2}{5} = 2$.

(c) We show the probabilities don't multiply:

$P(X = -2, Y = 0) = 0 \neq P(X = -2) \cdot P(Y = 0) = 1/25$.

Since these are not equal X and Y are not independent. (It is obvious that X^2 is not independent of X.)

(d) Using the table from part (a) and the means computed in part (d) we get:

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{5}(-2)(4) + \frac{1}{5}(-1)(1) + \frac{1}{5}(0)(0) + \frac{1}{5}(1)(1) + \frac{1}{5}(2)(4) = 0.$$

Problem 39. **(a)** $F(a,b) = P(X \le a, Y \le b) = \displaystyle\int_0^a \int_0^b (x+y)\, dy\, dx$.

Inner integral: $\left. xy + \dfrac{y^2}{2} \right|_0^b = xb + \dfrac{b^2}{2}$. Outer integral: $\left. \dfrac{x^2}{2} b + \dfrac{b^2}{2} x \right|_0^a = \dfrac{a^2 b + ab^2}{2}$.

So $\boxed{F(x,y) = \dfrac{x^2 y + xy^2}{2}}$ and $\boxed{F(1,1) = 1.}$

(b) $f_X(x) = \displaystyle\int_0^1 f(x,y)\, dy = \int_0^1 (x+y)\, dy = \left. xy + \frac{y^2}{2} \right|_0^1 = \boxed{x + \frac{1}{2}.}$

By symmetry, $\boxed{f_Y(y) = y + 1/2.}$

(c) To see if they are independent we check if the joint density is the product of the marginal densities.

$f(x,y) = x + y, \quad f_X(x) \cdot f_Y(y) = (x + 1/2)(y + 1/2)$.

Since these are not equal, $\boxed{X \text{ and } Y \text{ are not independent.}}$

(d) $E(X) = \displaystyle\int_0^1 \int_0^1 x(x+y)\, dy\, dx = \int_0^1 \left[\left. x^2 y + x\frac{y^2}{2} \right|_0^1 \right] dx = \int_0^1 x^2 + \frac{x}{2}\, dx = \boxed{\frac{7}{12}.}$

(Or, using (b), $E(X) = \displaystyle\int_0^1 x f_X(x)\, dx = \int_0^1 x(x + 1/2)\, dx = 7/12$.)

By symmetry $\boxed{E(Y) = 7/12.}$

$E(X^2 + Y^2) = \displaystyle\int_0^1 \int_0^1 (x^2 + y^2)(x+y)\, dy\, dx = \boxed{\frac{5}{6}.}$

$E(XY) = \displaystyle\int_0^1 \int_0^1 xy(x+y)\, dy\, dx = \boxed{\frac{1}{3}.}$

$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \dfrac{1}{3} - \dfrac{49}{144} = \boxed{-\dfrac{1}{144}.}$

Problem 40.

Standardize:

$$\begin{aligned}
P\left(\sum_i X_i < 30\right) &= P\left(\frac{\frac{1}{n}\sum X_i - \mu}{\sigma/\sqrt{n}} < \frac{30/n - \mu}{\sigma/\sqrt{n}}\right) \\
&\approx P\left(Z < \frac{30/100 - 1/5}{1/30}\right) \quad \text{(by the central limit theorem)} \\
&= P(Z < 3) \\
&= 0.9987 \quad \text{(from the table of normal probabilities)}
\end{aligned}$$

Problem 41. If $p < .5$ your expected winnings on any bet is negative, if $p = .5$ it is 0, and if $p > .5$ is is positive. By making a lot of bets the minimum strategy will 'win' you close to the expected average. So if $p \le .5$ you should use the maximum strategy and if $p > .5$ you should use the minumum strategy.

Problem 42. Let X_j be the IQ of a randomly selected person. We are given $E(X_j) = 100$ and $\sigma_{X_j} = 15$.
Let $\overline{X}$ be the average of the IQ's of 100 randomly selected people. We have $(\overline{X}) = 100$ and $\sigma_{\overline{X}} = 15/\sqrt{100} = 1.5$.
The problem asks for $P(\overline{X} > 115)$. Standardizing we get $P(\overline{X} > 115) \approx P(Z > 10)$. $\boxed{\text{This is effectively 0.}}$

Problem 43.

- A certain town is served by two hospitals.
- Larger hospital: about 45 babies born each day.
- Smaller hospital about 15 babies born each day.
- For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys.

(a) Which hospital do you think recorded more such days?

(i) The larger hospital. (ii) The smaller hospital.
(iii) About the same (that is, within 5% of each other).

(b) Let L_i (resp., S_i) be the Bernoulli random variable which takes the value 1 if more than 60% of the babies born in the larger (resp., smaller) hospital on the i^{th} day were boys. Determine the distribution of L_i and of S_i.

(c) Let L (resp., S) be the number of days on which more than 60% of the babies born in the larger (resp., smaller) hospital were boys. What type of distribution do L and S have? Compute the expected value and variance in each case.

(d) Via the CLT, approximate the .84 quantile of L (resp., S). Would you like to revise your answer to part (a)?

(e) What is the correlation of L and S? What is the joint pmf of L and S? Visualize the region corresponding to the event $L > S$. Express $P(L > S)$ as a double sum. **(a)** When this question was asked in a study, the number of undergraduates who chose each option was 21, 21, and 55, respectively. This shows a lack of intuition for the relevance of sample size on deviation from the true mean (i.e., variance).

(b) The random variable X_L, giving the number of boys born in the larger hospital on day i, is governed by a Bin(45, .5) distribution. So L_i has a Ber(p_L) distribution with

$$p_L = P(X > 27) = \sum_{k=28}^{45} \binom{45}{k} .5^{45} \approx .068$$

Similarly, the random variable X_S, giving the number of boys born in the smaller hospital on day i, is governed by a Bin(15, .5) distribution. So S_i has a Ber(p_S) distribution with

$$p_S = P(X_S > 9) = \sum_{k=10}^{15} \binom{15}{k} .5^{15} \approx .151$$

We see that p_S is indeed greater than p_L, consistent with (ii).

(c) Note that $L = \sum_{i=1}^{365} L_i$ and $S = \sum_{i=1}^{365} S_i$. So L has a Bin(365, p_L) distribution and S has a Bin(365, p_S) distribution. Thus

$$\begin{aligned} E(L) &= 365p_L \approx 25 \\ E(S) &= 365p_S \approx 55 \\ \mathrm{Var}(L) &= 365p_L(1 - p_L) \approx 23 \\ \mathrm{Var}(S) &= 365p_S(1 - p_S) \approx 47 \end{aligned}$$

(d) mean + sd in each case:

For L, $q_{.84} \approx 25 + \sqrt{23}$.

For S, $q_{.84} \approx 55 + \sqrt{47}$.

(e) Since L and S are independent, their joint distribution is determined by multiplying their individual distributions. Both L and S are binomial with $n = 365$ and p_L and p_S computed above. Thus

$$p_{l,s}P(L = i \text{ and } S = j) = p(i,j) = \binom{365}{i} p_L^i (1 - p_L)^{365-i} \binom{365}{j} p_S^j (1 - p_S)^{365-j}$$

Thus

$$P(L > S) = \sum_{i=0}^{364} \sum_{j=i+1}^{365} p(i,j) \approx .0000916$$

(We used R to do the computations.)

Problem 44. We compute the data mean and variance $\bar{x} = 65$, $s^2 = 35.778$. The number of degrees of freedom is 9. We look up the *critical value* $t_{9,.025} = 2.262$ in the t-table The 95% confidence interval is

$$\left[\bar{x} - \frac{t_{9,0.025}s}{\sqrt{n}},\ \bar{x} + \frac{t_{9,0.025}s}{\sqrt{n}}\right] = \left[65 - 2.262\sqrt{3.5778},\ 65 + 2.262\sqrt{3.5778}\right] = [60.721,\ 69.279]$$

Problem 45. Suppose we have taken data $x_1, \ldots, x_n$ with mean $\bar{x}$. The 95% confidence interval for the mean is $\overline{x} \pm z_{0.025}\frac{\sigma}{\sqrt{n}}$. This has width $2\,z_{0.025}\frac{\sigma}{\sqrt{n}}$. Setting the width equal to 1 and substitituting values $z_{0.025} = 1.96$ and $\sigma = 5$ we get

$$2 \cdot 1.96\frac{5}{\sqrt{n}} = 1 \quad \Rightarrow \quad \sqrt{n} = 19.6.$$

So, $n = (19.6)^2 = \boxed{384.}$.

If we use our rule of thumb that $z_{0.025} = 2$ we have $\sqrt{n}/10 = 2 \ \Rightarrow\ n = 400$.

Problem 46. The rule-of-thumb is that a 95% confidence interval is $\bar{x} \pm 1/\sqrt{n}$. To be within 1% we need

$$\frac{1}{\sqrt{n}} = 0.01 \ \Rightarrow\ n = 10000.$$

Using $z_{0.025} = 1.96$ instead the 95% confidence interval is

$$\bar{x} \pm \frac{z_{0.025}}{2\sqrt{n}}.$$

To be within 1% we need

$$\frac{z_{0.025}}{2\sqrt{n}} = 0.01 \ \Rightarrow\ n = 9604.$$

Note, we are still using the standard Bernoulli approximation $\sigma \leq 1/2$.

Problem 47. The 90% confidence interval is $\overline{x} \pm z_{0.05} \cdot \frac{1}{2\sqrt{n}}$. Since $z_{0.05} = 1.64$ and $n = 400$ our confidence interval is

$$\overline{x} \pm 1.64 \cdot \frac{1}{40} \ = \ \overline{x} \pm 0.041$$

If this is entirely above 0.5 we have $\overline{x} - 0.041 > 0.5$, so $\overline{x} > 0.541$. Let T be the number out of 400 who prefer A. We have $\overline{x} = \frac{T}{400} > 0.541$, so $\boxed{T > 216}$.

Problem 48. A 95% confidence means about $5\% = 1/20$ will be wrong. You'd expect about 2 to be wrong.

With a probability $p = 0.05$ of being wrong, the number wrong follows a Binomial$(40, p)$ distribution. This has expected value 2, and standard deviation $\sqrt{40(0.05)(0.95)} = 1.38$. 10 wrong is (10-2)/1.38 = 5.8 standard deviations from the mean. This would be surprising.

Problem 49. We have $n = 20$ and $s^2 = 4.06^2$. If we fix a hypothesis for σ^2 we know

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

We used R to find the critical values. (Or use the χ^2 table.)

```
c025 = qchisq(0.975,19) = 32.852
c975 = qchisq(0.025,19) = 8.907
```

The 95% confidence interval for σ^2 is

$$\left[\frac{(n-1)\cdot s^2}{c_{0.025}}, \frac{(n-1)\cdot s^2}{c_{0.975}}\right] = \left[\frac{19\cdot 4.06^2}{32.852}, \frac{19\cdot 4.66^2}{8.907}\right] = [9.53,\ 35.16]$$

We can take square roots to find the 95% confidence interval for σ

$$[3.09,\ 5.93]$$

Problem 50. **(a)** The model is $y_i = a + bx_i + \varepsilon_i$, where ε_i is random error. We assume the errors are independent with mean 0 and the same variance for each i (homoscedastic).

The total error squared is

$$E^2 = \sum (y_i - a - bx_i)^2 = (1-a-b)^2 + (1-a-2b)^2 + (3-a-3b)^2$$

The least squares fit is given by the values of a and b which minimize E^2. We solve for them by setting the partial derivatives of E^2 with respect to a and b to 0. In R we found that $a = 1.0$, $b = 0.5$

Also see the exam 2 and post exam 2 practice material and the practice final.