18.05 Problem Set 4, Spring 2014 Solutions

Problem 1. (10 pts.) **(a)**

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

(b) For $t \geq 0$, we know that $T \geq t$ if and only if both $X_1 \geq t$ and $X_2 \geq t$. So $P(T \geq t) = P(X_1 \geq t, X_2 \geq t)$. Since X_1 and X_2 are independent,

$$P(X_1 \geq t, X_2 \geq t) = P(X_1 \geq t)P(X_2 \geq t) = e^{-2\lambda t}.$$

Thus, $F_T(t) = P(T \leq t) = 1 - e^{-2\lambda t}$. Differentiating with respect to t to get the pdf, we find

$$f_T(t) = F_T'(t) = 2\lambda e^{-2\lambda t}.$$

T is an exponential random variable with mean $\frac{1}{2\lambda}$.

(c) Let X_1, X_2, and X_3 be the lifetimes of bulbs B_1, B_2 and B_3, respectively. Then we know $X_1 \sim \exp(2), X_2 \sim \exp(3), X_3 \sim \exp(5)$. Let $T = \min(X_1, X_2, X_3)$. Then T is the time to the first failure of a bulb. Following the same argument as in (b), we have

$$P(T \geq t) = P(X_1 \geq t)P(X_2 \geq t)P(X_3 \geq t) = \mathrm{e}^{-10t}.$$

Thus, the cdf of T is $F_T(t) = 1 - e^{-10t}$ and the pdf, $f_T(t)$ is given by

$$f_T(t) = F_T'(t) = 10e^{-10t}$$

and we find that $T \sim \exp(10)$. Therefore, $E[T] = \frac{1}{10}$.

Problem 2. (10 pts.) **(a)** We have

$$1 = \int_0^1 \int_0^1 c(x^2 + xy) dy dx = c \int_0^1 x^2 + \frac{x}{2} dx = c(\frac{1}{3} + \frac{1}{4}) = \frac{7c}{12}.$$

Thus, $\boxed{c = \frac{12}{7}}$. We have

$$\begin{aligned} F(x,y) = P(X \leq x, Y \leq y) &= \frac{12}{7} \int_0^x \int_0^y u^2 + uv\, dy\, dx \\ &= \frac{12}{7} \int_0^x u^2 y + \frac{uy^2}{2}\, du \\ &= \boxed{\frac{12}{7}\left(\frac{x^3 y}{3} + \frac{x^2 y^2}{4}\right)}. \end{aligned}$$

(b) The marginal pdf's are:

$$\begin{aligned} f_X(x) &= \int_0^1 f(x,y)\, dy = \boxed{\frac{12}{7}\left(x^2 + \frac{x}{2}\right)} \\ f_Y(y) &= \int_0^1 f(x,y)\, dx = \boxed{\frac{12}{7}\left(\frac{1}{3} + \frac{y}{2}\right)}. \end{aligned}$$

The marginal cdf's are:

$$F_X(x) = F(x, 1) = \boxed{\frac{12}{7}\left(\frac{x^3}{3} + \frac{x^2}{4}\right)}$$
$$F_Y(y) = F(1, y) = \boxed{\frac{12}{7}\left(\frac{y}{3} + \frac{y^2}{4}\right)}.$$

(c)

$$E(X) = \int_0^1 x f_X(x)\,dx = \frac{12}{7}\int_0^1 x\left(x^2 + \frac{x}{2}\right)dx = \frac{12}{7}\left(\frac{1}{4} + \frac{1}{6}\right) = \boxed{\frac{5}{7} \approx 0.7143}$$
$$E(X^2) = \int_0^1 x^2 f_X(x)\,dx = \frac{39}{70} \approx 0.5571.$$

Thus $\boxed{\mathrm{Var}(X) = E(X^2) - E(X)^2 \approx 0.0469}$.

(d)

$$E(Y) = \int_0^1 y f_Y(y)\,dy = \frac{12}{7}\int_0^1 y\left(\frac{1}{3} + \frac{y}{2}\right)dy = \frac{4}{7} \approx 0.5714$$
$$E(Y^2) = \int_0^1 y^2 f_Y(y) = \frac{12}{7}\int_0^1 y^2\left(\frac{1}{3} + \frac{y}{2}\right)dy = \frac{17}{42} \approx 0.4048$$
$$\mathrm{Var}(Y) = E(Y^2) - E(Y)^2 \approx 0.0782$$
$$E(XY) = \int_0^1\int_0^1 xy f(x, y)\,dy\,dx = \frac{12}{7}\int_0^1\int_0^1 x^3 y + x^2 y^2\,dy\,dx = \frac{17}{42} \approx 0.4048$$
$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) \approx \boxed{-0.0034}$$
$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \boxed{-0.0561}$$

Problem 3. (10 pts.) **(a)** Define

$$X_i = \begin{cases} 1 & \text{if person } i \text{ supports Erika} \\ 0 & \text{if person } i \text{ does not support Erika} \end{cases}$$

Then $X_i \sim \mathrm{Bern}(0.5)$ and the number of people who prefer Erika is

$$S = X_1 + \cdots + X_{400}.$$

We know $E(X_i) = 1/2$ and $\mathrm{Var}(X_i) = 1/4$. This implies $E(S) = 200$ and $\mathrm{Var}(S) = 100$. Thus the central limit theorem tells us that

$$S \approx \mathrm{N}(200, 100).$$

The problem asks for $P(S > 210)$:

$$P(S > 210) = P\left(\frac{S - 200}{10} > \frac{210 - 200}{10}\right) \approx P(Z > 1) \approx \boxed{0.16}.$$

(b) If we now let $Y_i = 1$ if person i prefers one of Peter, Jon or Jerry and 0 otherwise, we have $Y_1, \ldots, Y_{400}$ independent Bern(0.3). So $E(Y_i) = \mu = 0.3$ and $\text{Var}(Y_i) = (0.3)(0.7) = 0.21$. If $\overline{Y}_{400} = \frac{1}{400}(Y_1 + \cdots + Y_{400})$, the Central Limit Theorem tells us

$$\frac{\overline{Y}_n - \mu}{\sigma/\sqrt{400}} = \frac{(\overline{Y}_n - 0.3)\sqrt{400}}{\sqrt{0.21}}$$

is approximately standard normal. If Z is standard normal, then

$$P(\overline{Y} \leq 0.25) \approx P\left(Z < \frac{(0.25 - 0.3)\sqrt{400}}{\sqrt{0.21}}\right) \approx 0.0145.$$

Problem 4. (10 pts.)

Let S be the total rounding error for a day. The problems asks for

$$P(|S| > 100).$$

Let X_i be the rounding error (in cents) of the i^{th} order. Then X_i takes values $-2, -1, 0, 1, 2$, each with probability $\frac{1}{5}$. We compute

$$E(X_i) = \mu = 0, \quad \text{Var}(X_i) = \sigma^2 = 2.$$

The total rounding error $S = X_1 + \cdots + X_{1000}$. By the Central Limit Theorem, we know that $S \approx \text{N}(0, 2000)$.

$$P(|S| \geq 100) = P(\left(\frac{|S - 0|}{\sqrt{2000}} \geq \frac{100}{\sqrt{2000}}\right) \approx P\left(|Z| \geq \frac{100}{\sqrt{2000}}\right) = \boxed{0.02534}.$$

Extra credit 5 points Here's my code.

```
r = c(0,-1,-2,2,1) # rounding error for 0, 1, 2, 3, 4 cents
ntrials = 10000
data = rep(0,1000)
for (j in 1:ntrials)
{
   x = sample(r,1000,replace=TRUE) #rounding from 1000 orders
   trial = sum(x) # total rounding error
   data[j] = trial
}
mean(abs(data) > 100 ) # fraction of rounding errors > 100 or < -100
```

In three runs it gave 0.0269, 0.0265, 0.0276. This agrees nicely with the CLT estimate.

Problem 5. (10 pts.) From the table we compute the marginal probabilities

$$P(X = 1) = \frac{1}{3}, \quad P(Y = 1) = \frac{1}{3}.$$

Since $P(X = 1, Y = 1) = \frac{1}{18}$ and $P(X = 1)P(Y = 1) = \frac{1}{9}$, X and Y are not independent.

Problem 6. (10 pts.) Solution: **(a)** The marginal distributions allow us to determine the joint distribution of X and Y in terms of c:

$Y \backslash X$	1	-1	
1	c	$.5-c$	.5
-1	$.5-c$	c	.5
	.5	.5	

We easily compute: $E(X) = 0$, $E(Y) = 0$, $\text{Var}(X) = 1$, $\text{Var}(Y) = 1$. So, computing directly

$$\begin{aligned} E(XY) &= (1 \cdot 1)c + (-1 \cdot 1)(.5-c) + (1 \cdot -1)(.5-c) + (-1 \cdot -1)c \\ &= 4c - 1 \end{aligned}$$

Thus,

$$\begin{aligned} \text{Cor}(X,Y) &= E(XY) - E(X)E(Y) = 4c - 1 \\ \text{Cov}(X,Y) &= \frac{\text{Cor}(X,Y)}{\sigma_X \sigma_Y} = 4c - 1. \end{aligned}$$

(b) Note that the correlation runs from -1 to 1 as c runs from 0 to .5.

We must have $\text{Cov}(X,Y) = 0$ for X and Y to be independent. This only happens when $c = \frac{1}{4}$. It is easy to check in this case that all four probabilities in the table are 0.5 and they are independent.

When $c = 0$ the correlation is -1, which means X and Y are fully correlated (sometimes called fully anti-correlated). When $c = 0.5$ the correlation is 1.0 and X and Y are fully correlated.

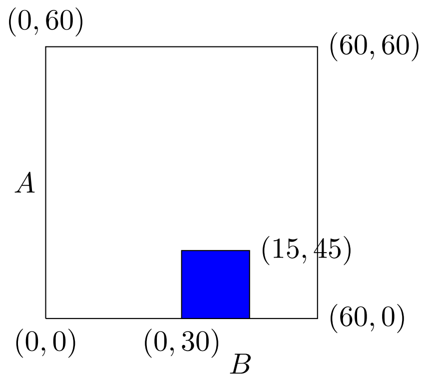
Problem 7. (10 pts.) **(a)** The joint probability density function is $f(a,b) = \frac{1}{3600}$ and the joint cumulative density function is

$$F(a,b) = \int_0^a \int_0^b f(s,t)\, ds\, dt = \frac{ab}{3600}$$

(b) Since A is uniformly distributed on $[0,60]$, $P(A \le 30) = \frac{1}{2}$.

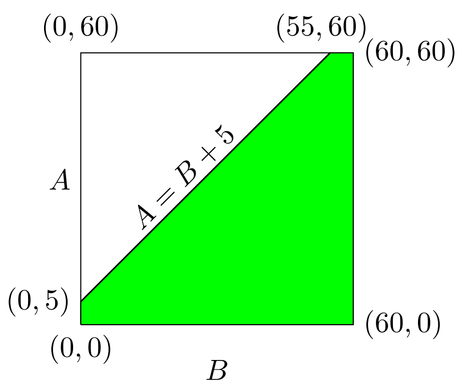
(c) i) $P(A \le 15, 30 \le B \le 45) = P(A \le 15)P(30 \le B \le 45) = 0.0625$

ii) The range of (A,B) is the square $[0,60] \times [0,60]$. The event 'Alice arrives before 12:15 and Bob arrives between 12:30 and 12:45' is represented by the solid blue rectangle. Since the probability distribution is uniform the probability of the blue rectangle is just the fraction of the entire square the it covers.



(d) The shaded area in the figure below corresponds to the event '$A \leq B+5$'. (Note: if Alice arrives before Bob then she arrives less than 5 minutes after him.) That is, it corresponds to all pairs of arrival times (a, b) such that $a \leq b+5$. $P(A \leq B+5)$ is then just the area of the green region divided by the area of the entire square. We find that the area of the white region is $\frac{55^2}{2}$. So

$$P(A \leq B+5) = \frac{1}{3600}\left(3600 - \frac{55^2}{2}\right) = \boxed{0.5799}.$$

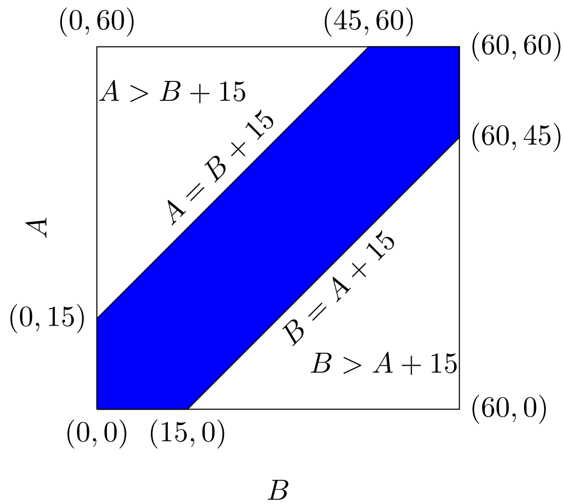


(e) Alice and Bob arrive within 15 minutes of each other is event

$$E = \text{`}B - 15 \leq A \leq B + 15'.$$

This is the blue shaded region in the figure below. We see that the area of each white triangle is $\frac{45^2}{2}$. So, the combined white area is 45^2 and

$$P(E) = \frac{3600 - 45^2}{3600} = \boxed{\frac{7}{16}}.$$


(0, 60)
(45, 60)
(60, 60)
A > B + 15
A = B + 15
(60, 45)
A
B = A + 15
(0, 15)
B > A + 15
(60, 0)
(0, 0)
(15, 0)
B