DATA 605 Week 5 Homework

Ilya Kats

October 1, 2017

Problem

Choose independently two numbers B and C at random from the interval $[0, 1]$ with uniform density. Prove that B and C are proper probability distributions.

Find the probability that

(a) $B + C < 1/2$
(b) $BC < 1/2$
(c) $|B - C| < 1/2$
(d) $max\{B, C\} < 1/2$
(e) $min\{B, C\} < 1/2$

Solution

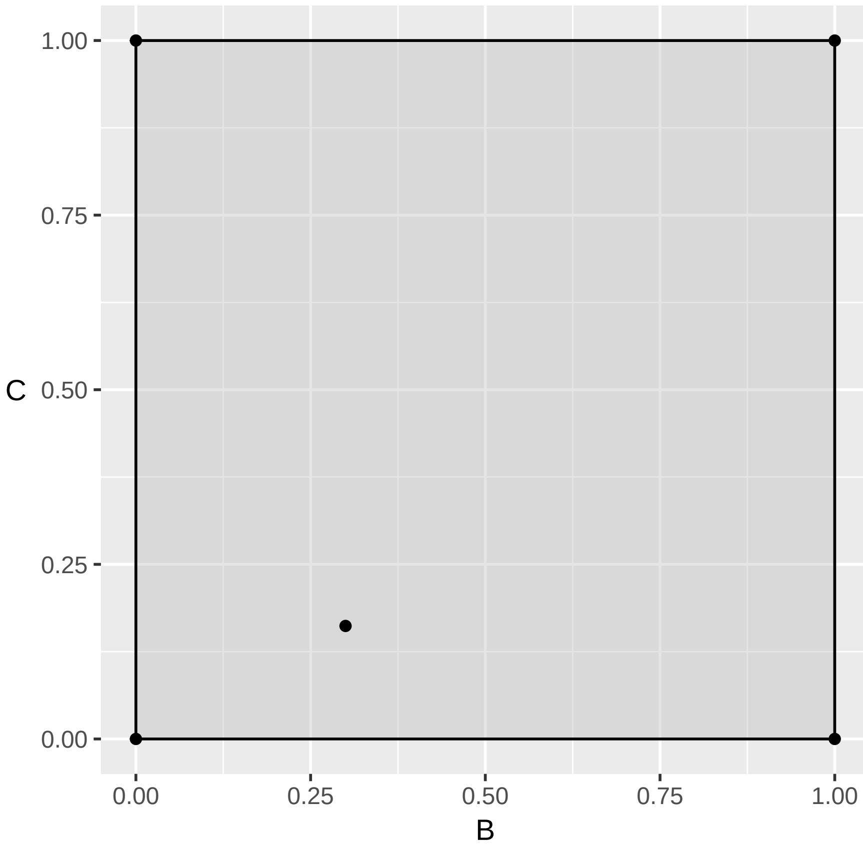
Since B and C are selected with uniform density, let us define the density function as $f(x) = \begin{cases} 1, & \text{if } 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$

Then $P(0 \le X \le 1) = \int_0^1 f(x)dx = 1$.

The value of $f(x)$ is greater than or equal to 0 for all x and total area under the density function is equal to 1. As such this is a proper probability distribution that satisfies both B and C.

Now, consider a unit square with a randomly chosen point (B, C).

```
library(ggplot2)
ggplot()+
  geom_rect(aes(xmin=0, xmax=1, ymin=0,ymax=1), fill="grey", alpha=0.4, color="black")+
  geom_point(aes(c(1,0,0,1,runif(1,0,1)),c(1,0,1,0,runif(1,0,1))))+
  xlim(0,1)+ylim(0,1)+coord_fixed()+
  xlab("B")+ylab("C")+
  theme(axis.title.y = element_text(angle = 0, vjust=0.5))
```



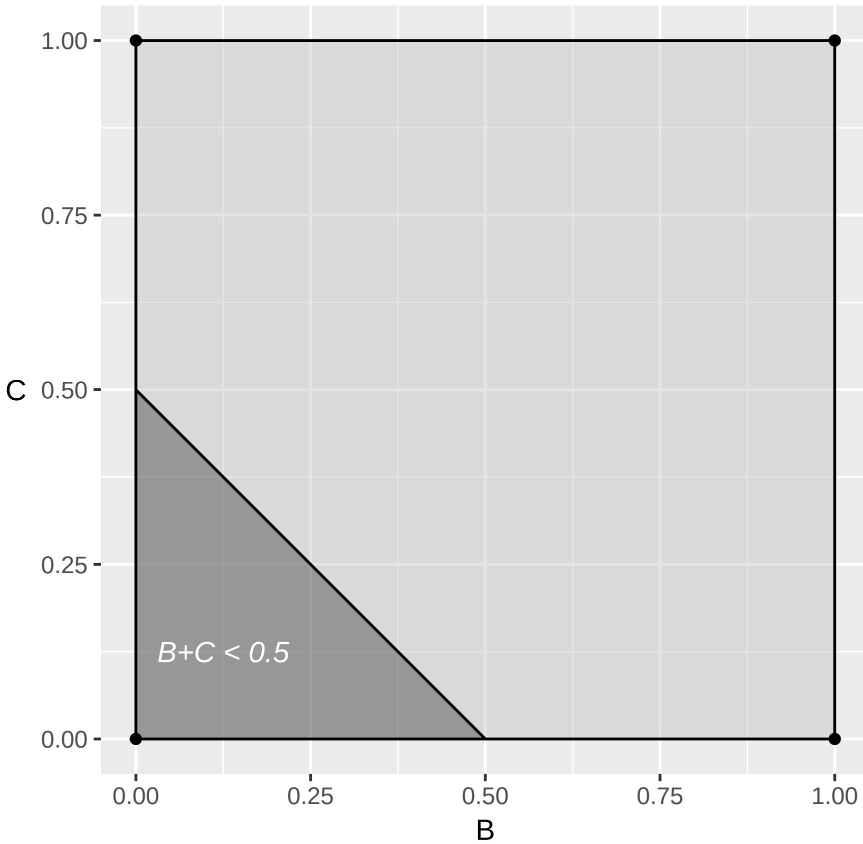
Part (a): $B + C < 1/2$

Consider $x + y = 0.5$, where x represents B and y represents C. Then $y = 0.5 - x$. If we plot this line in the unit square, then the area under the line will be all values of B and C such that $B + C < 0.5$ and the area will equal the probability $P(B + C < 0.5)$.

```
x <- seq(from=0,to=0.5,length.out=1000)
y <- 0.5-x

# Define polygon for under the curve shading
shade <- rbind(c(0,0), data.frame(x,y))

ggplot()+
  geom_rect(aes(xmin=0, xmax=1, ymin=0,ymax=1), fill="grey", alpha=0.4, color="black")+
  geom_point(aes(c(1,0,0,1),c(1,0,1,0)))+
  xlim(0,1)+ylim(0,1)+coord_fixed()+
  xlab("B")+ylab("C")+
  theme(axis.title.y = element_text(angle = 0, vjust=0.5))+
  geom_line(aes(x,y))+
  geom_polygon(aes(shade$x,shade$y), fill="black", alpha=0.3)+
  geom_text(aes(0.125,0.125), label="B+C < 0.5", size=4, color="white", fontface="italic")
```



$P(B+C<0.5)=\frac{0.5^2}{2}=\frac{0.25}{2}=0.125$

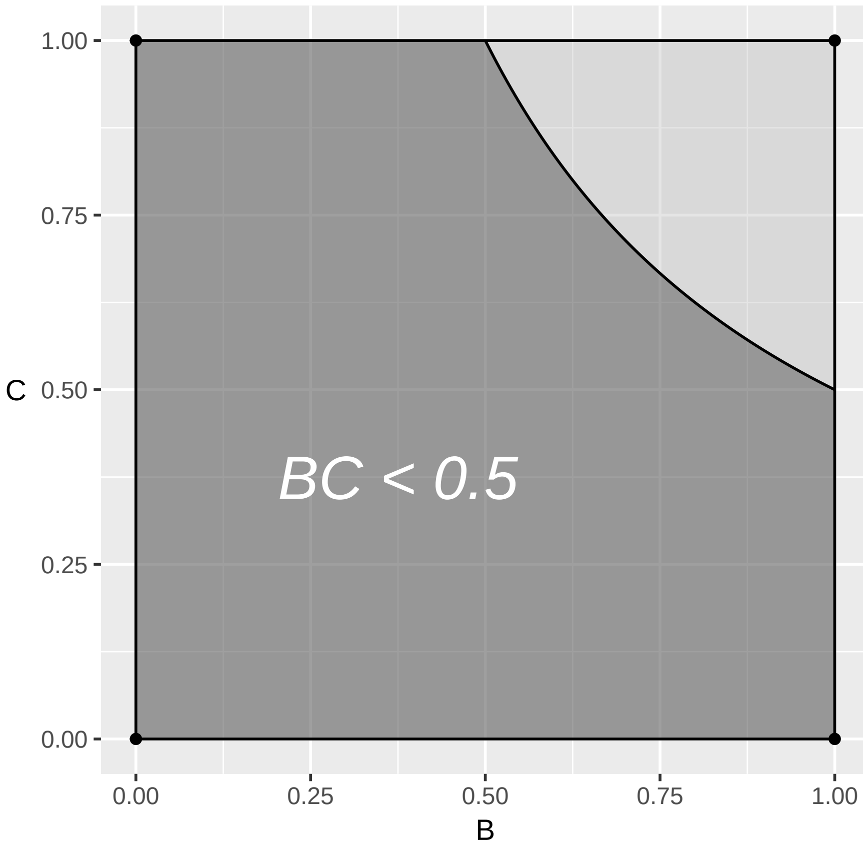
Part (b): $BC<1/2$

Consider $xy=0.5$, where x represents B and y represents C. Then $y=\frac{0.5}{x}=\frac{1}{2x}$. If we plot this line in the unit square, then the area under the line will be all values of B and C such that $BC<0.5$ and the area will equal the probability $P(BC<0.5)$.

```
x <- seq(from=0.5,to=1,length.out=1000)
y <- 1/(2*x)

# Define polygon for under the curve shading
shade <- rbind(c(0,0), c(0,1), c(0.5,1), data.frame(x,y), c(1, 0))

ggplot()+
  geom_rect(aes(xmin=0, xmax=1, ymin=0,ymax=1), fill="grey", alpha=0.4, color="black")+
  geom_point(aes(c(1,0,0,1),c(1,0,1,0)))+
  xlim(0,1)+ylim(0,1)+coord_fixed()+
  xlab("B")+ylab("C")+
  theme(axis.title.y = element_text(angle = 0, vjust=0.5))+
  geom_line(aes(x,y))+
  geom_polygon(aes(shade$x,shade$y), fill="black", alpha=0.3)+
  geom_text(aes(0.375,0.375), label="BC < 0.5", size=8, color="white", fontface="italic")
```



We can integrate to get the area under the curve and add 0.5 of the left half of the unit square.

$P(BC < 0.5) = 0.5 + \int_{0.5}^{1} \frac{1}{2x} dx = 0.5 + 0.346574 = 0.846574$

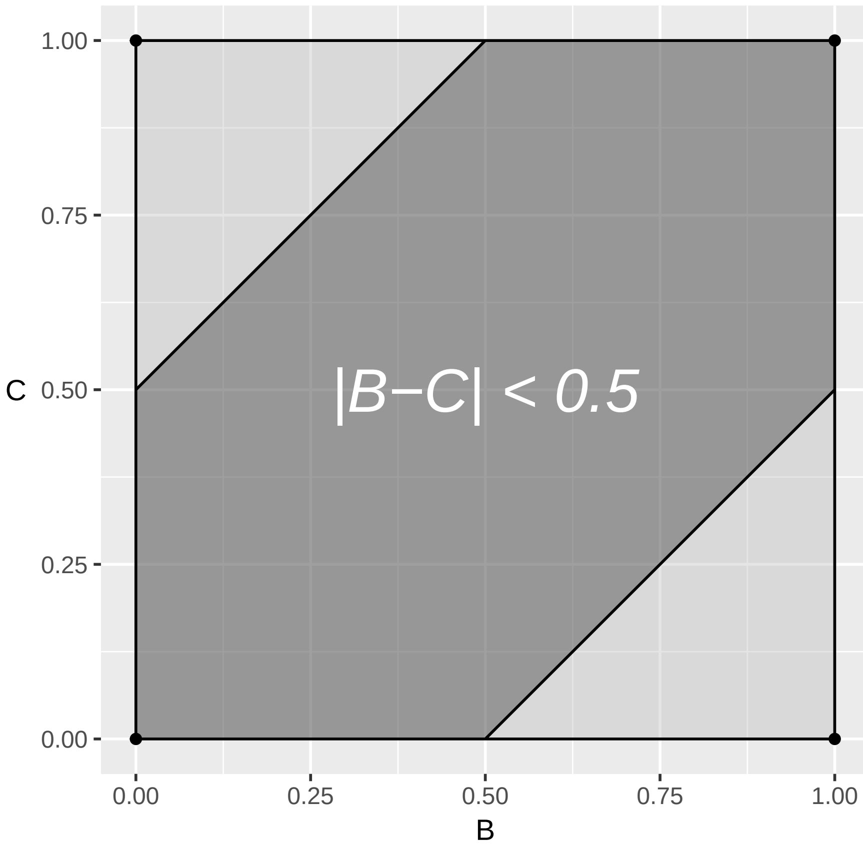
Part (c): $|B - C| < 1/2$

Similarly to the above consider two lines - $x - y = 0.5$ and $x - y = -0.5$.

```
x1 <- seq(from=0.5,to=1,length.out=2)
x2 <- seq(from=0,to=0.5,length.out=2)
y1 <- x1-0.5
y2 <- x2+0.5

# Define polygon for under the curve shading
shade <- cbind(c(0,0,0.5,1,1,0.5), c(0,0.5,1,1,0.5,0))

ggplot()+
  geom_rect(aes(xmin=0, xmax=1, ymin=0,ymax=1), fill="grey", alpha=0.4, color="black")+
  geom_point(aes(c(1,0,0,1),c(1,0,1,0)))+
  xlim(0,1)+ylim(0,1)+coord_fixed()+
  xlab("B")+ylab("C")+
  theme(axis.title.y = element_text(angle = 0, vjust=0.5))+
  geom_line(aes(x1,y1))+
  geom_line(aes(x2,y2))+
  geom_polygon(aes(shade[,1],shade[,2]), fill="black", alpha=0.3)+
  geom_text(aes(0.5,0.5), label="|B-C| < 0.5", size=8, color="white", fontface="italic")
```



We can easily see that the probability of the event is 1 minus two triangles similar to part (a).

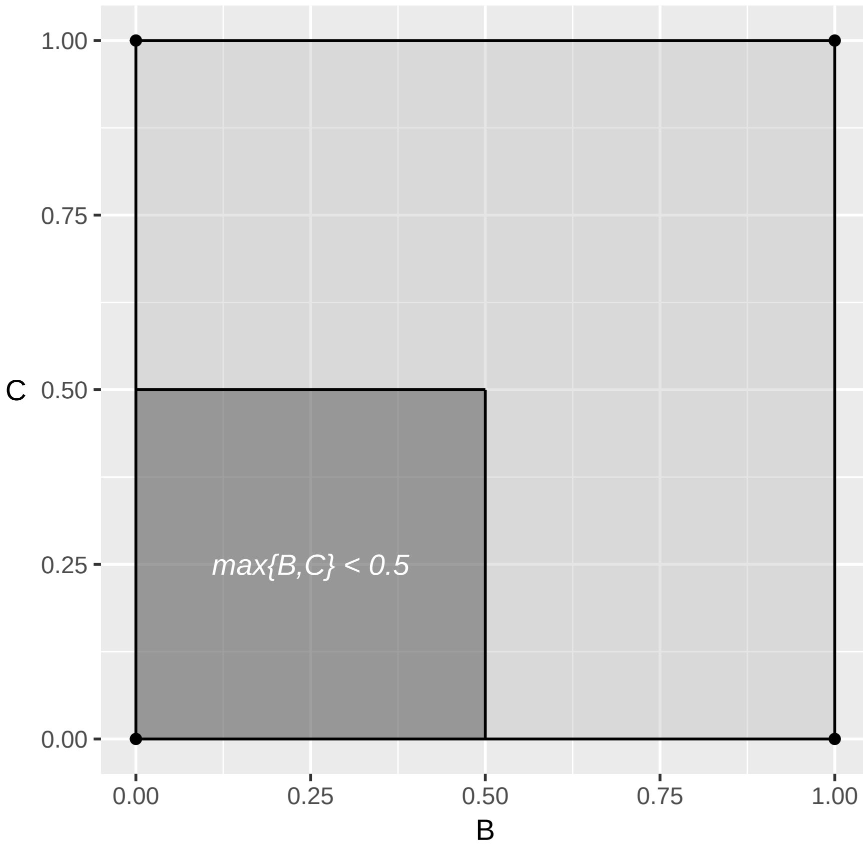
$P(|B - C| < 0.5) = 1 - 2 \times 0.125 = 0.75$

Part (d): $max\{B, C\} < 1/2$

Any combination of B and C such that $B < 0.5$ and $C < 0.5$, will satisfy $max\{B, C\} < 0.5$. If $B > 0.5$, then either $B > C$ and $max\{B, C\} = B > 0.5$ or $0.5 < B < C$ and $max\{B, C\} = C > 0.5$. Similarly for $C > 0.5$.

```
# Define polygon for under the curve shading
shade <- cbind(c(0,0,0.5,0.5), c(0,0.5,0.5,0))

ggplot()+
  geom_rect(aes(xmin=0, xmax=1, ymin=0,ymax=1), fill="grey", alpha=0.4, color="black")+
  geom_point(aes(c(1,0,0,1),c(1,0,1,0)))+
  xlim(0,1)+ylim(0,1)+coord_fixed()+
  xlab("B")+ylab("C")+
  geom_line(aes(c(0,0.5),c(0.5,0.5)))+
  geom_line(aes(c(0.5,0.5),c(0,0.5)))+
  theme(axis.title.y = element_text(angle = 0, vjust=0.5))+
  geom_polygon(aes(shade[,1],shade[,2]), fill="black", alpha=0.3)+
  geom_text(aes(0.25,0.25), label="max{B,C} < 0.5", size=4, color="white", fontface="italic")
```



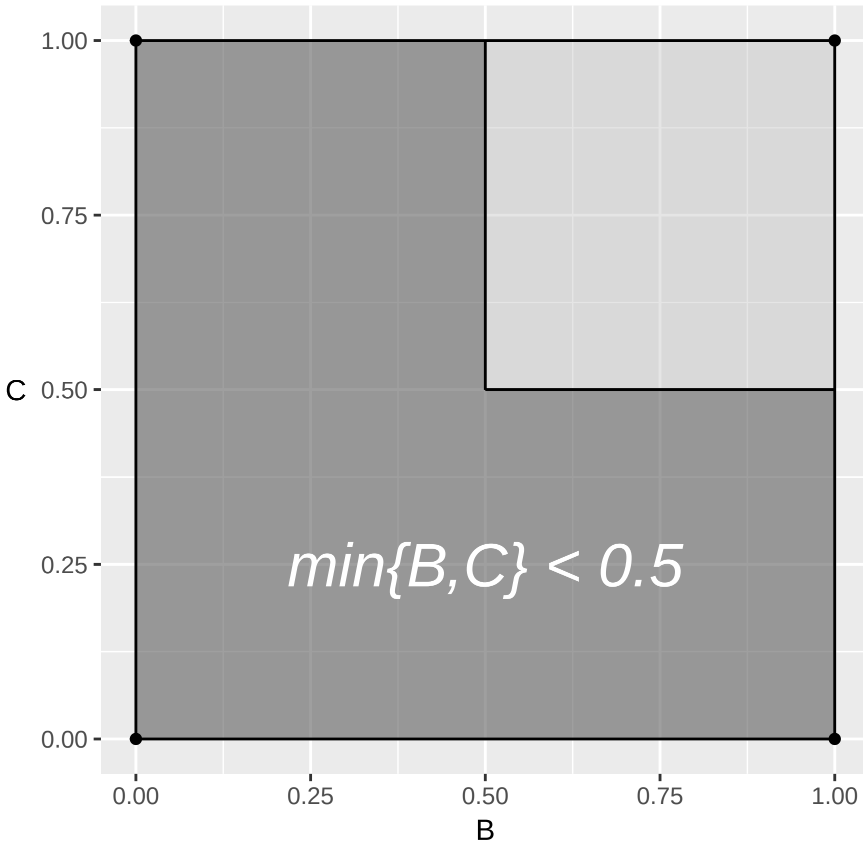
$P(max\{B,C\} < 0.5) = 0.25$

Part (e): $min\{B,C\} < 1/2$

Any combination of B and C such that $B < 0.5$ or $C < 0.5$, will satisfy $min\{B,C\} < 0.5$. It is only if both B and C are greater than 0.5, then $min\{B,C\} > 0.5$.

```
# Define polygon for under the curve shading
shade <- cbind(c(0,0,0.5,0.5,1,1), c(0,1,1,0.5,0.5,0))

ggplot()+
  geom_rect(aes(xmin=0, xmax=1, ymin=0,ymax=1), fill="grey", alpha=0.4, color="black")+
  geom_point(aes(c(1,0,0,1),c(1,0,1,0)))+
  xlim(0,1)+ylim(0,1)+coord_fixed()+
  xlab("B")+ylab("C")+
  geom_line(aes(c(0.5,0.5),c(0.5,1)))+
  geom_line(aes(c(0.5,1),c(0.5,0.5)))+
  theme(axis.title.y = element_text(angle = 0, vjust=0.5))+
  geom_polygon(aes(shade[,1],shade[,2]), fill="black", alpha=0.3)+
  geom_text(aes(0.5,0.25), label="min{B,C} < 0.5", size=8, color="white", fontface="italic")
```



$P(min\{B,C\} < 0.5) = 0.75$

Answer

(a) $P(B + C < 1/2) = 0.125$
(b) $P(BC < 1/2) = 0.846574$
(c) $P(|B - C| < 1/2) = 0.75$
(d) $P(max\{B,C\} < 1/2) = 0.25$
(e) $P(min\{B,C\} < 1/2) = 0.75$

Simulation

Double-check results via simulation. Modified from code posted on BlackBoard.

```r
n<-1000000
B<-runif(n,0,1)
C<-runif(n,0,1)

partA<-((B+C)<0.5)

partB<-((B*C)<0.5)

partC<-(abs(B-C)<0.5)

partD<-rep(0, n)
partE<-rep(0, n)
```

```r
for (i in 1:n) {
  partD[i]<-max(B[i],C[i])
  partE[i]<-min(B[i],C[i])
}
partD<-(partD<0.5)
partE<-(partE<0.5)

simulation <- cbind(c("B+C<0.5", "BC<0.5", "|B-C|<0.5", "max(B,C)<0.5", "min(B,C)<0.5"),
                    c(sum(partA)/n,sum(partB)/n,sum(partC)/n,sum(partD)/n,sum(partE)/n)
                    )
colnames(simulation) <- c("Event", "Probability")
rownames(simulation) <- c("a","b","c","d","e")
simulation
```

```
##   Event          Probability
## a "B+C<0.5"      "0.124185"
## b "BC<0.5"       "0.846469"
## c "|B-C|<0.5"    "0.749773"
## d "max(B,C)<0.5" "0.249492"
## e "min(B,C)<0.5" "0.749743"
```