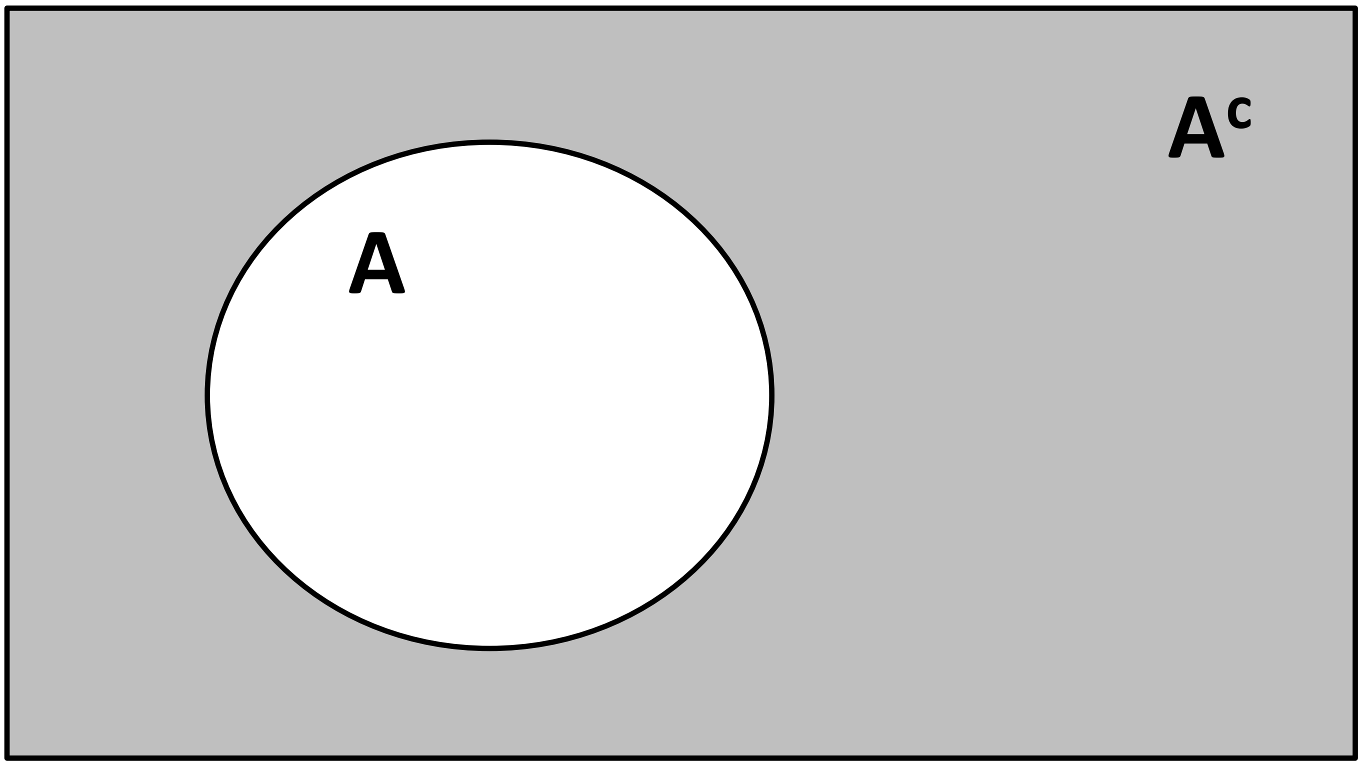
確率

「ハーバード大学講義テキスト 生物統計学入門」６章

- 参考書4&5章は寿命データ「生存時間解析(21章)」に特化するので、ここでは省略
- ベイズ統計や条件付き確率を駆使した診断テストの計算を目標

確率事象に関するベン図

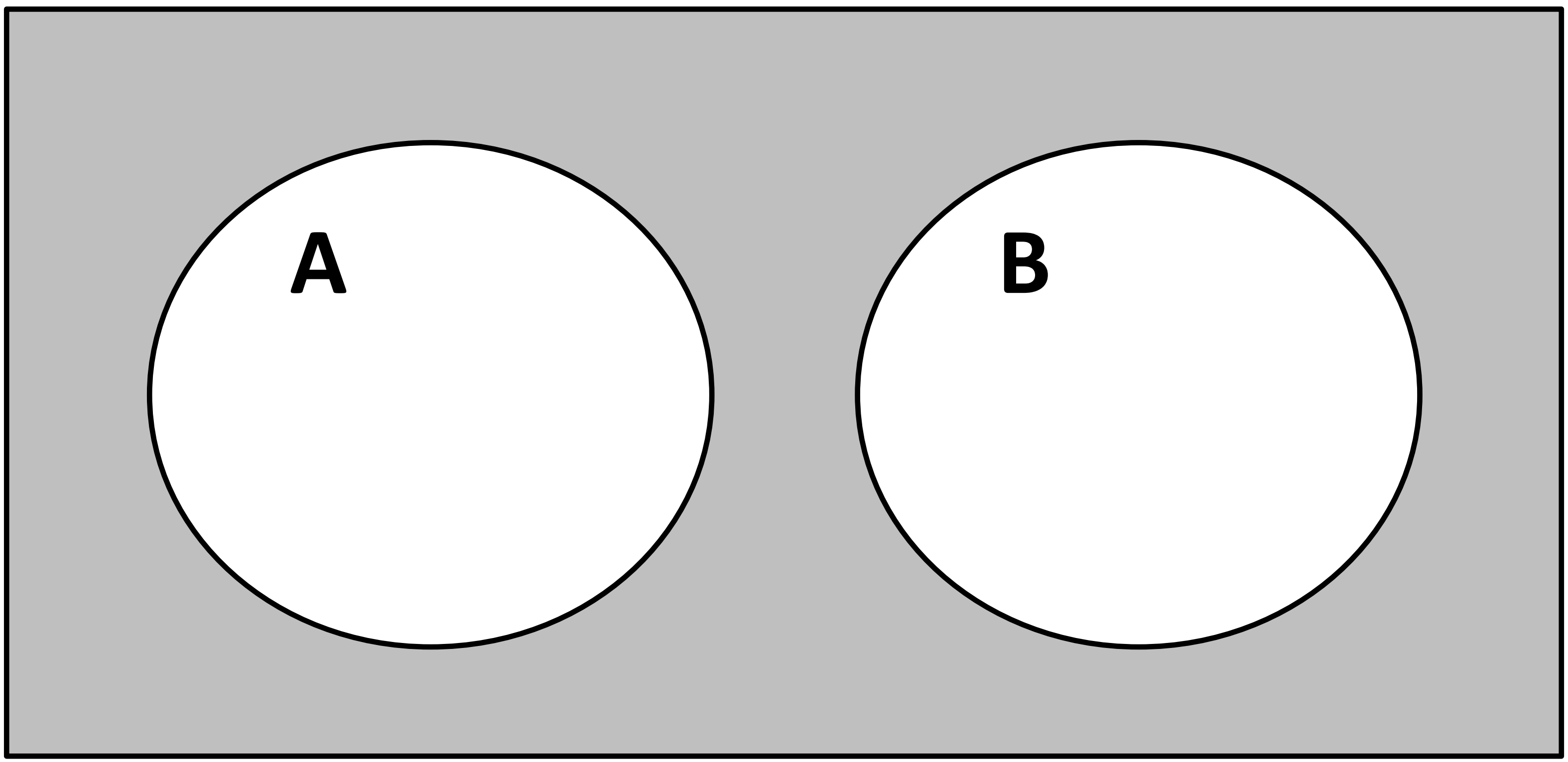


例） A:「乳児が最初の１年間生存する」

$P(A)=0.99149$, $P(A^c)=1-P(A)=0.00851$,

$P(A \cup A^c)=1$, $P(A \cap A^c)=0$.

排反事象



例）　A:「乳児の体重が2000g未満」

B:「乳児の体重が2000～2499g」

$P(A \cap B)=0$,　 $A \cap B=\emptyset$ (空), $P(A \cup B)=P(A)+P(B)=0.025+0.043=0.068$

排反事象の例

1992年に米国で出産した女性の年齢

年　齢	確　率
＜15	0.003
15-19	0.124
20-24	0.263
24-29	0.290
30-34	0.220
35-39	0.085
40-44	0.014
45-49	0.001
計	1.000

表 7.1　米国で生まれた子供の出生順の確率変数 X の確率分布

x	$P(X=x)$
1	0.416
2	0.330
3	0.158
4	0.058
5	0.021
6	0.009
7	0.004
8^+	0.004
計	1.000

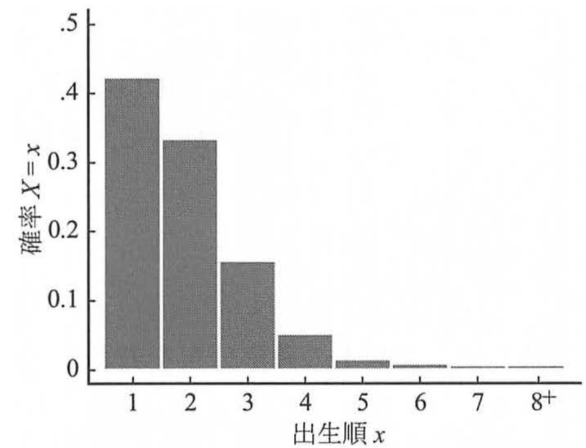
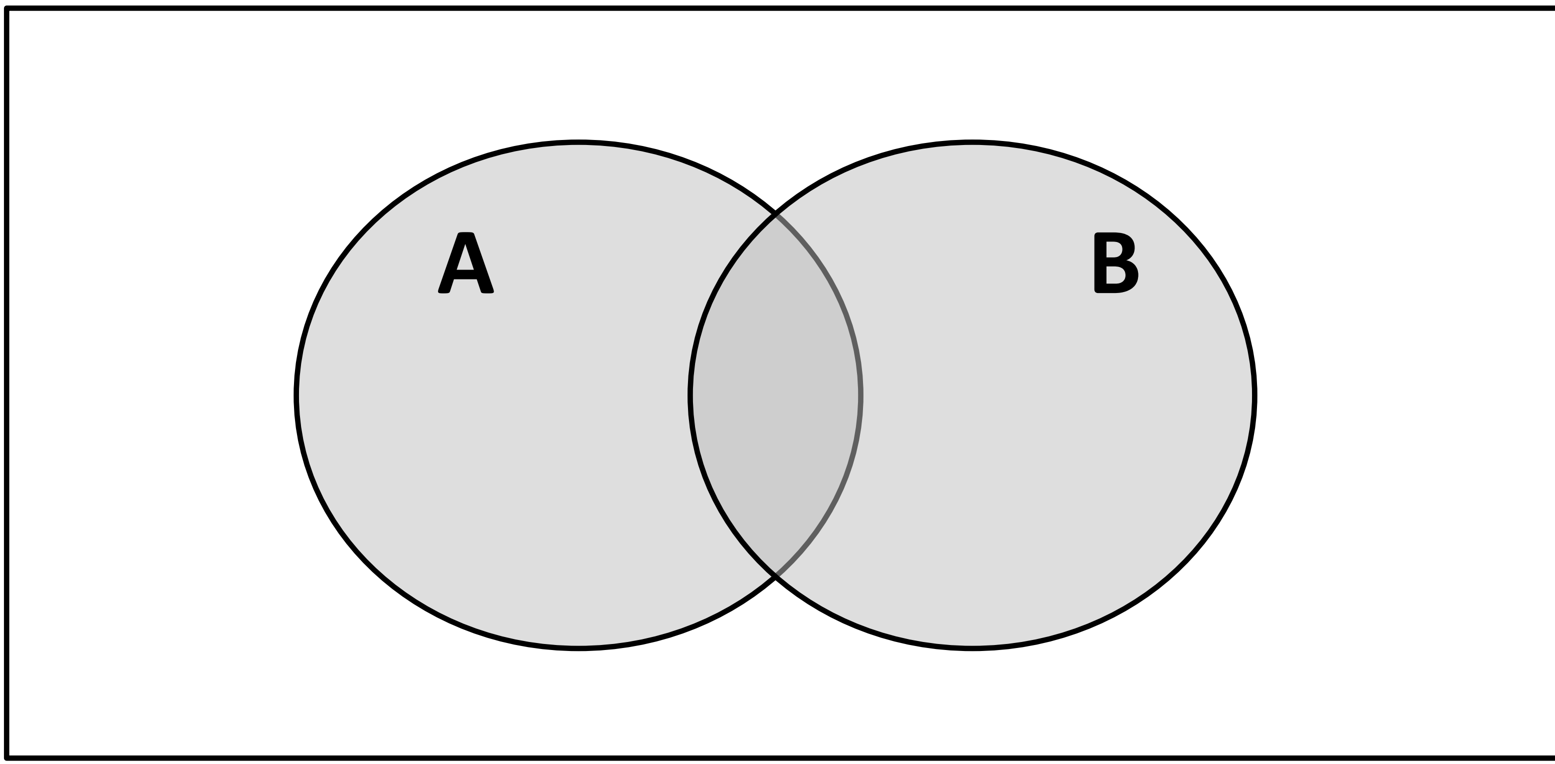


図 7.1　米国で生まれた子供の出生順の確率変数 X の確率分布

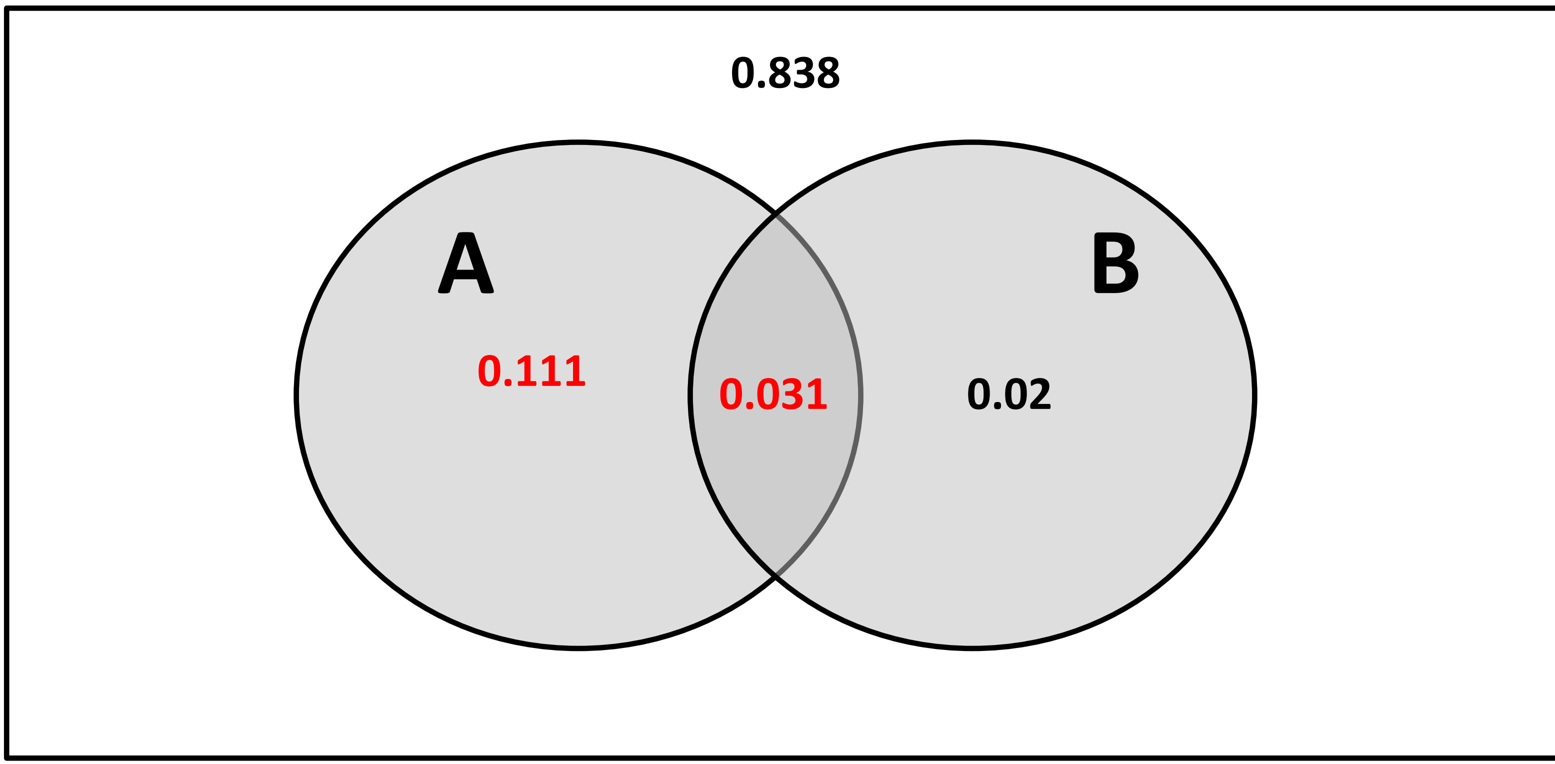
交わり（intersection）



例） A :「新生児の妊娠期間が37週未満」
B :「新生児の体重が2500gより軽い」
P(A)=0.142， P(B)=0.051， P(A∩B) = 0.031

条件付き確率: P(B|A)

新生児の妊娠期間が37週未満の時、体重が2500gより軽い確率は？

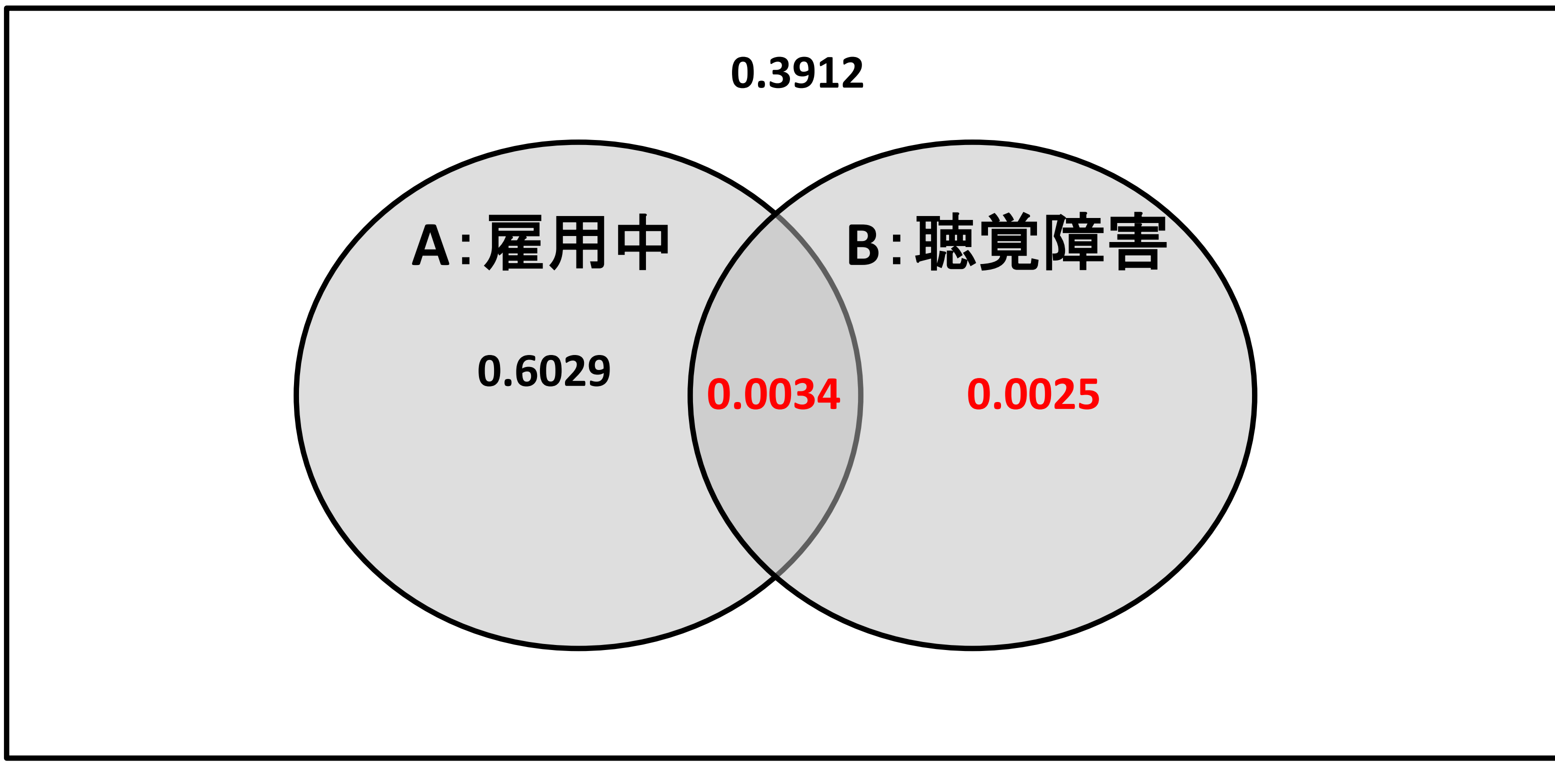


例） P(B|A) = **0.031** / (**0.111**+**0.031**) = 0.218

一般には P(B|A) = P(A∩B) / P(A)

∵ P(B|A) P(A) = P(A∩B)

条件付き確率: P(A|B)



聴覚障害者が雇用されている条件付き確率を求めよ。

$P(A|B) = P(A \cap B) / P(B) = \mathbf{0.0034} / (\mathbf{0.0034} + \mathbf{0.0025}) = 0.5763$

ベイズの定理の応用

$$\boxed{P(B|A)\,P(A) \;=\; P(A \cap B) \;=\; P(A|B)\,P(B)}$$

不確実なパップ塗布による子宮癌のスクリーニング:
P(陰性|癌患者) = 0.1625
P(陽性|癌患者) = 0.8375
P(陰性|癌患者でない) = 0.8136
P(陽性|癌患者でない) = 0.1864

また、別の統計データより
P(癌患者) = 0.000083

ベイズの定理により、
陽性が出た場合に実際に癌である確率を求める:
P(癌患者|陽性) = P(陽性|癌患者) P(癌患者) / P(陽性) = **0.000373**
ただしここで、以下を用いた。
P(陽性) = P(陽性|癌患者) P(癌患者) + P(陽性|癌患者でない) P(癌患者でない)
= 0.1865

正確性に問題があることがわかった。

ベイズの定理の応用2

$$P(B|A)\,P(A) = P(A \cap B) = P(A|B)\,P(B)$$

不確実な胸部X線による結核のスクリーニング:
P(陰性|結核) = 0.2667
P(陽性|結核) = 0.7333
P(陰性|結核でない) = 0.9715
P(陽性|結核でない) = 0.0285

また、別の統計データより
P(結核) = 0.000093

ベイズの定理により、
陽性が出た場合に実際に結核である確率を求める:
P(結核|陽性) = P(陽性|結核) P(結核) / P(陽性) = **0.00239**
ただしここで、以下を用いた。
P(陽性) = P(陽性|結核) P(結核) + P(陽性|結核でない) P(結核でない) = 0.0286

結核である確率は、陽性であることにより 25.7倍上昇する

ベイズの定理の応用3

$$P(B|A)\,P(A) = P(A \cap B) = P(A|B)\,P(B)$$

不確実な放射性脳室造影による冠状動脈疾患のスクリーニング:

P(陰性|疾患有) = 0.372
P(陽性|疾患有) = 0.628
P(陰性|疾患無) = 0.823
P(陽性|疾患無) = 0.177

また、別の統計データより
P(疾患有) = 0.1

ベイズの定理により、
陽性が出た場合に実際に疾患有である確率を求める:
P(疾患有|陽性) = P(陽性|疾患有) P(疾患有) / P(陽性)
= 0.628 * 0.1 / 0.222 = **0.283**
ただしここで、以下を用いた。
P(陽性) = P(陽性|疾患有) P(疾患有) + P(陽性|疾患無) P(疾患無)
= 0.628 * 0.1 + 0.177 * 0.9 = 0.222

ベイズの定理の応用4 P(感染者|陽性)

$$P(B|A)\,P(A) \;=\; P(A\cap B) \;=\; P(A|B)\,P(B)$$

不確実なD社の抗体検査による新型コロナのスクリーニング：

P(陰性|感染者) = 0.2
P(陽性|感染者) = 0.8
P(陰性|感染者でない) = 1
P(陽性|感染者でない) = 0

日本感染症学会2020年4月17日
http://www.kansensho.or.jp/uploads/files/news/gakkai/covid19_kensakit_0423.pdf

また、別の統計データより
P(感染者) = 0.001

ベイズの定理により、
「陽性が出た場合に実際に感染者である」確率を求める：
P(感染者|陽性) = P(陽性|感染者) P(感染者) / P(陽性) = **1**
ただしここで、以下を用いた。
P(陽性) = P(陽性|感染者) P(感染者) + P(陽性|感染者でない) P(感染者でない)
= 0.0008

偽陽性の問題は無いことがわかった。

ベイズの定理の応用4 P(感染者|陰性)

$$\boxed{P(B|A)\,P(A) = P(A \cap B) = P(A|B)\,P(B)}$$

不確実なD社の抗体検査による新型コロナのスクリーニング：
P(陰性|感染者) = 0.2
P(陽性|感染者) = 0.8
P(陰性|感染者でない) = 1
P(陽性|感染者でない) = 0

日本感染症学会2020年4月17日
http://www.kansensho.or.jp/uploads/files/news/gakkai/covid19_kensakit_0423.pdf

また、別の統計データより
P(感染者) = 0.001

ベイズの定理により、
「陰性が出た場合でも実際には感染者である」確率を求める：
P(感染者|陰性) = P(陰性|感染者) P(感染者) / P(陰性) = **0.0002**
ただしここで、以下を用いた。
P(陰性) = P(陰性|感染者) P(感染者) + P(陰性|感染者でない) P(感染者でない)
= 0.9992

P(感染者|陰性)は著しく小さく、P(感染者)のさらに0.2倍だとわかる。