Conditional Probability

e.g. 1

- A bag has 5 balls
- 3 are black
- 2 are white
- Draw a ball, and replace – do it 4 times
- Probability of getting 2 black balls

No. of ways we can get 2 black balls

{B, B, B, B}	{W, B, B, B}
{B, B, B, W}	**{W, B, B, W}**
{B, B, W, B}	**{W, B, W, B}**
{B, B, W, W}	{W, B, W, W}
{B, W, B, B}	**{W, W, B, B}**
{B, W, B, W}	{W, W, B, W}
{B, W, W, B}	{W, W, W, B}
{B, W, W, W}	{W, W, W, W}

6 possible ways – what is the probability?

$$\frac{6}{16} = \frac{3}{8} \text{ (Is it correct?)}$$

What is the probability of this string?

{B, B, B, B}	{W, B, B, B}
{B, B, B, W}	**{W, B, B, W}**
{B, B, W, B}	**{W, B, W, B}**
{B, B, W, W}	{W, B, W, W}
{B, W, B, B}	**{W, W, B, B}**
{B, W, B, W}	{W, W, B, W}
{B, W, W, B}	{W, W, W, B}
{B, W, W, W}	{W, W, W, W}

$$\frac{3}{5}.\frac{3}{5}.\frac{2}{5}.\frac{2}{5} = \frac{36}{625} \rightarrow 6.\frac{36}{625}$$

Generalize

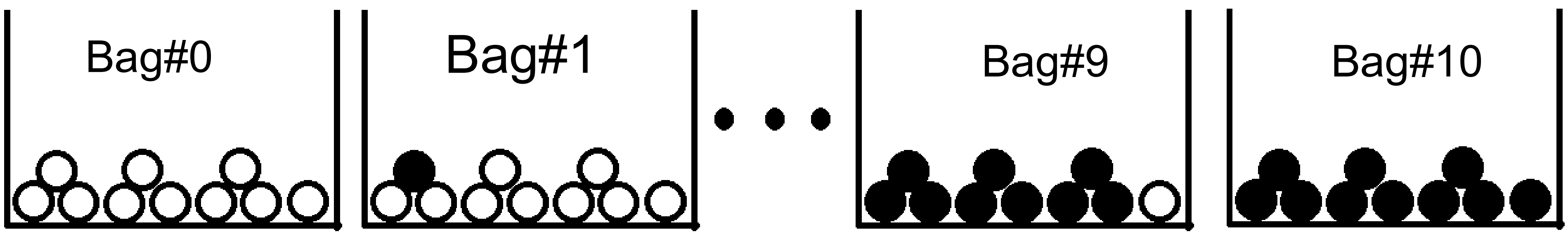
- A bag has K balls
- B are black
- W are white = (K-B)
- Draw a ball, and replace – do it N times
- Probability of getting n_B black balls

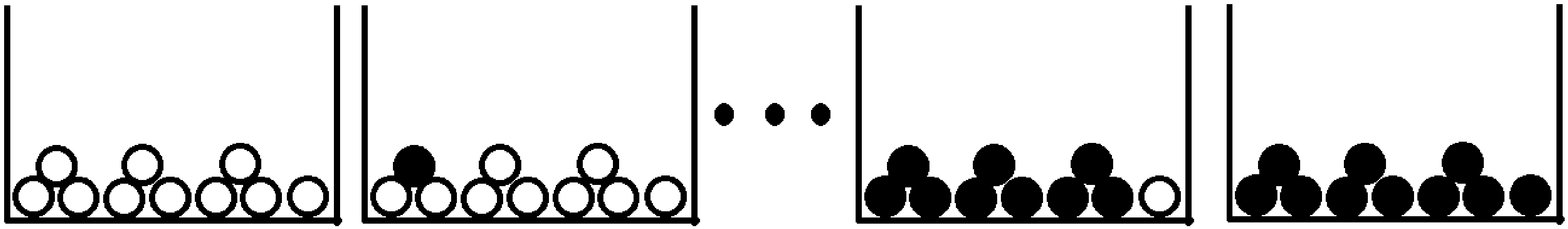
$$\text{define the fraction } f_B \equiv B/K$$

$$P(n_B \mid f_B, N) = \binom{N}{n_B} f_B^{n_B} (1 - f_B)^{N - n_B}$$

e.g.2

- 11 bags named from 0 to 10
- Each has 10 balls
- Bag#0 has 0 black balls, remaining white balls
- Bag#1 has 1 black ball, remaining white balls
- …
- Bag#10 has 10 black balls, remaining white balls





- Randomly choose a bag
- N times take a ball from the chosen bag with replacement
- Say we got 3 blacks in 10 times
- Can you say which bag is being used?

Answer

- No definite answer
- Guess work
- Every bag has certain probability values
- Bag#0 and bag#10 are not used– why?

Conditional probability – generalization

$$P(A \cap B) = P(\mathrm{A}|B).\mathrm{P(B)}$$

$$P(A \cap B \cap C) = P(\mathrm{A}|(\mathrm{B} \cap \mathrm{C}).\mathrm{P}(\mathrm{B} \cap \mathrm{C})$$

$$= P(\mathrm{A}|(\mathrm{B} \cap \mathrm{C}).\mathrm{P(B|C)}P(C)$$

For $n = 4$, i.e. four events, the chain rule reads

$$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4)$$
$$= \mathbb{P}(A_4 \mid A_3 \cap A_2 \cap A_1)\mathbb{P}(A_3 \cap A_2 \cap A_1)$$
$$= \mathbb{P}(A_4 \mid A_3 \cap A_2 \cap A_1)\mathbb{P}(A_3 \mid A_2 \cap A_1)\mathbb{P}(A_2 \cap A_1)$$
$$= \mathbb{P}(A_4 \mid A_3 \cap A_2 \cap A_1)\mathbb{P}(A_3 \mid A_2 \cap A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_1)$$

Conditional probability – Chain rule

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathbb{P}(A_n \mid A_1 \cap \ldots \cap A_{n-1})\,\mathbb{P}(A_1 \cap \ldots \cap A_{n-1})$$

Again apply the same logic

$$\mathbb{P}(A_1 \cap \ldots \cap A_{n-1}) = \mathbb{P}(A_{n-1} \mid A_1 \cap \ldots \cap A_{n-2})\,\mathbb{P}(A_1 \cap \ldots \cap A_{n-2})$$

$$\begin{aligned}
&\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) \\
&= \mathbb{P}(A_n \mid A_1 \cap \ldots \cap A_{n-1})\,\mathbb{P}(A_{n-1} \mid A_1 \cap \ldots \cap A_{n-2})\,\mathbb{P}(A_1 \cap \ldots \cap A_{n-2}) \\
&= \mathbb{P}(A_n \mid A_1 \cap \ldots \cap A_{n-1})\,\mathbb{P}(A_{n-1} \mid A_1 \cap \ldots \cap A_{n-2}) \cdot \ldots \cdot \mathbb{P}(A_3 \mid A_1 \cap A_2)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_1) \\
&= \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2) \cdot \ldots \cdot \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1})
\end{aligned}$$

e.g.

- One of the useful forms of the conditional probability
- LHS - Say it in words
- If H has happens then what is the probability that x and y happens

$$P(x, y \mid \mathcal{H}) = P(x \mid y, \mathcal{H}) P(y \mid \mathcal{H}) = P(y \mid x, \mathcal{H}) P(x \mid \mathcal{H})$$

$$P(x, y \mid \mathcal{H}) = \boxed{P(x \mid y, \mathcal{H})} P(y \mid \mathcal{H}) = P(y \mid x, \mathcal{H}) P(x \mid \mathcal{H})$$

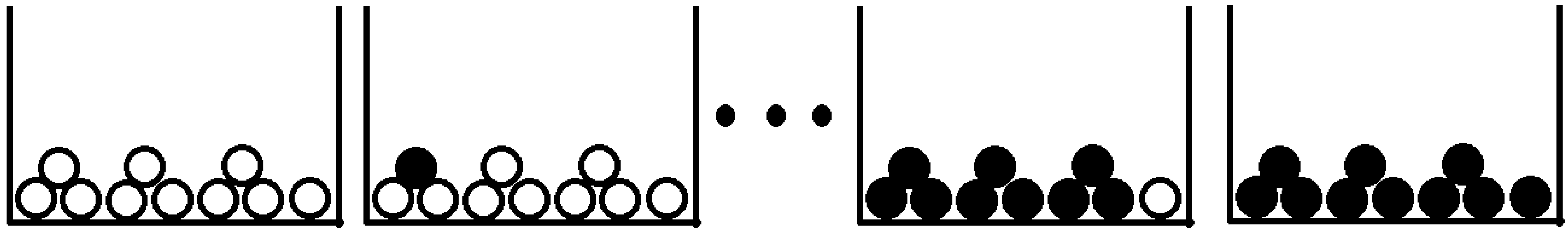
$$P(x|y, H) = \frac{P(x, y, H)}{P(y, H)}$$

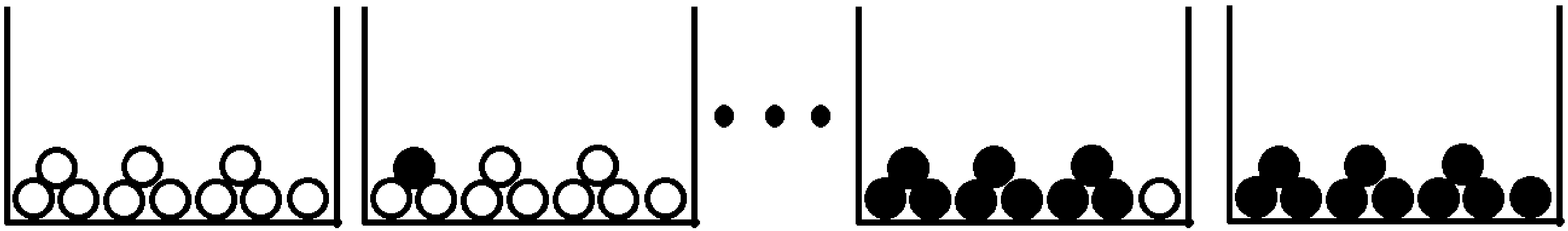
$$P(x, y|H) = \frac{P(x,y,H)}{P(y,H)} P(y|H) = \frac{P(x,y,H)}{P(y,H)} \frac{P(y,H)}{P(H)}$$

$$= \frac{P(x,y,H)}{P(H)} = \frac{P(x,y|H).P(H)}{P(H)} = P(x,y/H)$$

e.g.

- 11 bags named from 0 to 10
- Each has 10 balls
- Bag#0 has 0 black balls, remaining white balls
- Bag#1 has 1 black ball, remaining white balls
- …
- Bag#10 has 10 black balls, remaining white balls





- Randomly choose a bag
- N times take a ball from the chosen bag with replacement
- Say we got 3 blacks in 10 times
- Can you say which bag is being used?

Make use of conditional probability

- Given: 10 times we played and got 3 blacks
- To find: Bag#...
- Let u be bag#

i.e. $u \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

- N – number of times we play the game
- n_B – number of black balls

Approach

- We need to find, $\mathbf{P(u|n_B,N)}$
- If n_B and N are fixed, then u is the only variable
  - **For various values of u find the conditional probability**
- We do not know $\mathbf{P(u|n_B,N)}$
- But we can find out, $P(n_B|u,N)$ – refer e.g#1

Recall,

$$P(x, y \mid \mathcal{H}) = P(x \mid y, \mathcal{H}) P(y \mid \mathcal{H}) = P(y \mid x, \mathcal{H}) P(x \mid \mathcal{H})$$

$$P\,(u, n_B | N) = P(u | n_B, N) P(n_B | N) = P(n_B | u, N) P(u | N)$$

$$P\,(u, n_B|N) = P(u|n_B, N)P(n_B|N) = P(n_B|u, N)P(u|N)$$

$$P(u|n_B, N) = \frac{P(u, n_B|N)}{P(n_B|N)} = \frac{P(n_B|u, N).\,P(u|N)}{P(n_B|N)}$$

With any chosen u, we would have played N times i.e. they are independent

$$P(u|n_B, N) = \frac{P(n_B|u, N).\,P(u)}{P(n_B|N)}$$

$$P(u) = \frac{1}{11}$$

Let, $f_u \equiv u/10$

$$P(n_B\,|\,u, N) = \binom{N}{n_B} f_u^{n_B}(1 - f_u)^{N-n_B}$$

Let, $f_u \equiv u/10$

$$P(n_B \mid u, N) = \binom{N}{n_B} f_u^{n_B} (1 - f_u)^{N - n_B}$$

$$P(u|n_B, N) = \frac{\boxed{P(n_B|u, N)} . \boxed{P(u)}}{\boxed{P(n_B|N)}}$$

$$P(u) = \frac{1}{11}$$

$$P(n_B \mid N) = \sum_u P(u, n_B \mid N) = \sum_u P(u) P(n_B \mid u, N)$$

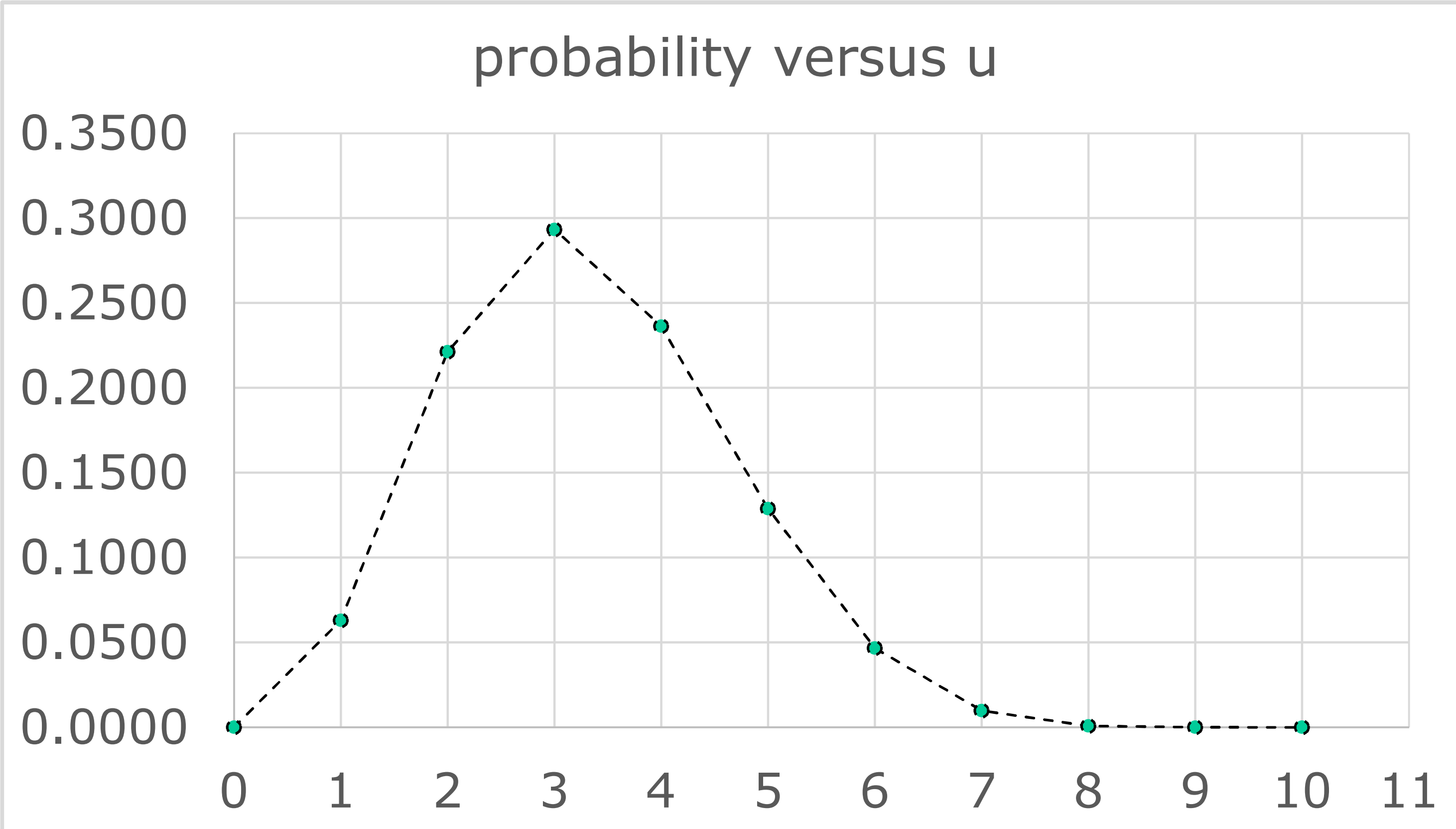
Different u gives different values – changes f_u

$$P(n_B = 3|u = 0, N = 10) = \binom{10}{3}\left(\frac{0}{10}\right)^3 . \left(1 - \frac{0}{10}\right)^7 = 0$$

$$P(n_B = 3|u = 1, N = 10) = \binom{10}{3}\left(\frac{1}{10}\right)^3 . \left(1 - \frac{1}{10}\right)^7$$

$$= \frac{10!}{7!\,.3!} . \frac{1}{1000} . \frac{9^7}{10^7} = 0.24\ X\ 0.478 = 0.115$$

N	nB	u	fu=u/10	P(nb|u,N)	P(u)	P(nb|u,N)*P(u)	
10	3	0	0.00	0.0000	0.0909	0.0000	0.0000
10	3	1	0.10	0.0574	0.0909	0.0052	0.0631
10	3	2	0.20	0.2013	0.0909	0.0183	0.2213
10	3	3	0.30	0.2668	0.0909	0.0243	0.2933
10	3	4	0.40	0.2150	0.0909	0.0195	0.2363
10	3	5	0.50	0.1172	0.0909	0.0107	0.1288
10	3	6	0.60	0.0425	0.0909	0.0039	0.0467
10	3	7	0.70	0.0090	0.0909	0.0008	0.0099
10	3	8	0.80	0.0008	0.0909	0.0001	0.0009
10	3	9	0.90	0.0000	0.0909	0.0000	0.0000
10	3	10	1.00	0.0000	0.0909	0.0000	0.0000
						0.0827	



In 10 drawings, we have got 3 black balls.
Probability that we could have chosen bag#3 is around 0.3

Change N = 20

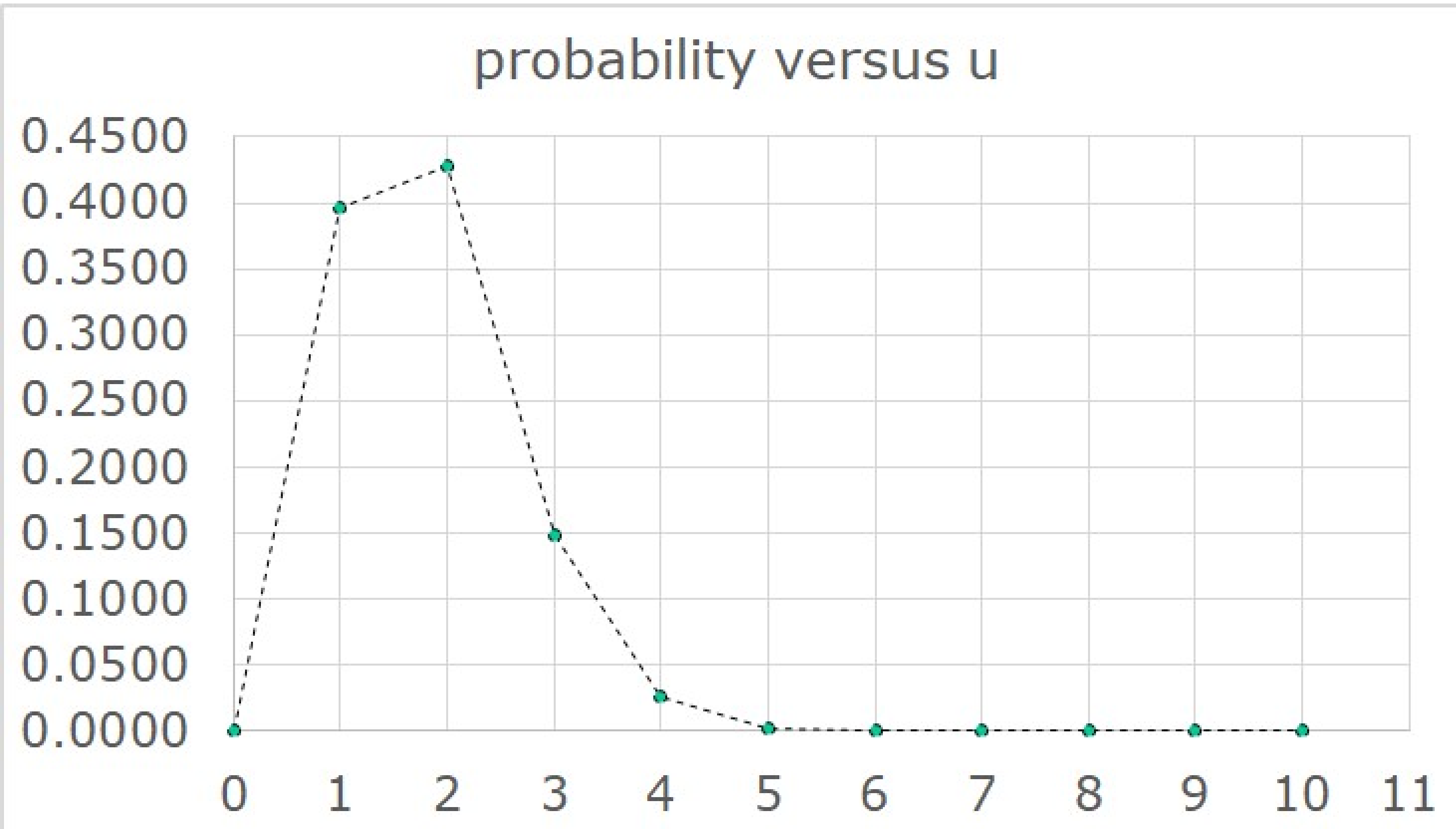
- Given: 20 times we played and got 3 blacks
- To find: Bag#...
- Let u be bag#

i.e. $u \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

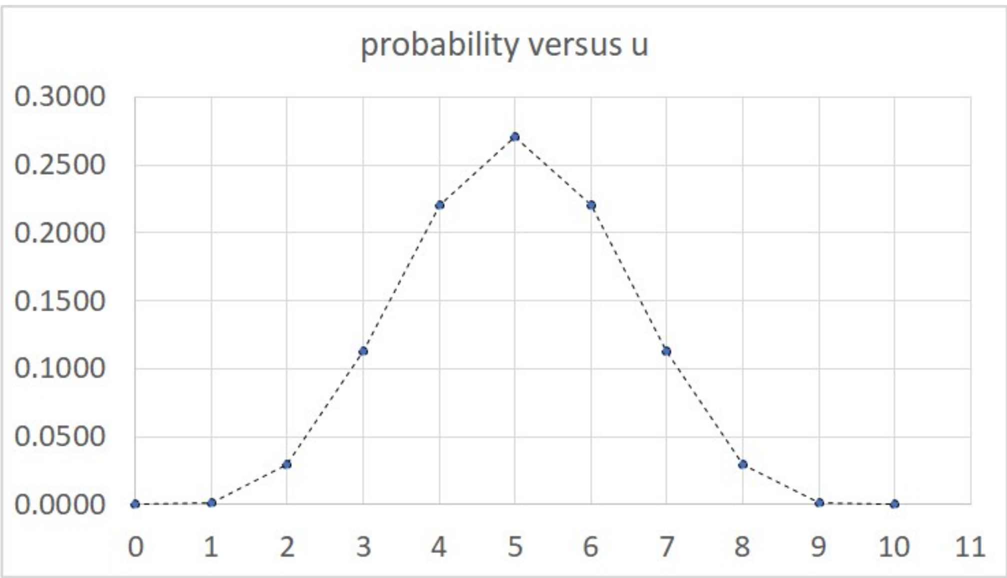
- N – number of times we play the game
- n_B – number of black balls

N	nB	u	fu=u/10	P(nb|u,N)	P(u)	P(nb|u,N)*P(u)	
20	3	0	0.00	0.0000	0.0909	0.0000	0.0000
20	3	1	0.10	0.1901	0.0909	0.0173	0.3955
20	3	2	0.20	0.2054	0.0909	0.0187	0.4272
20	3	3	0.30	0.0716	0.0909	0.0065	0.1490
20	3	4	0.40	0.0123	0.0909	0.0011	0.0257
20	3	5	0.50	0.0011	0.0909	0.0001	0.0023
20	3	6	0.60	0.0000	0.0909	0.0000	0.0001
20	3	7	0.70	0.0000	0.0909	0.0000	0.0000
20	3	8	0.80	0.0000	0.0909	0.0000	0.0000
20	3	9	0.90	0.0000	0.0909	0.0000	0.0000
20	3	10	1.00	0.0000	0.0909	0.0000	0.0000
						0.0437	

In 20 drawings, we have got 3 black balls.
Probability that we could have chosen bag#2 is around 0.43

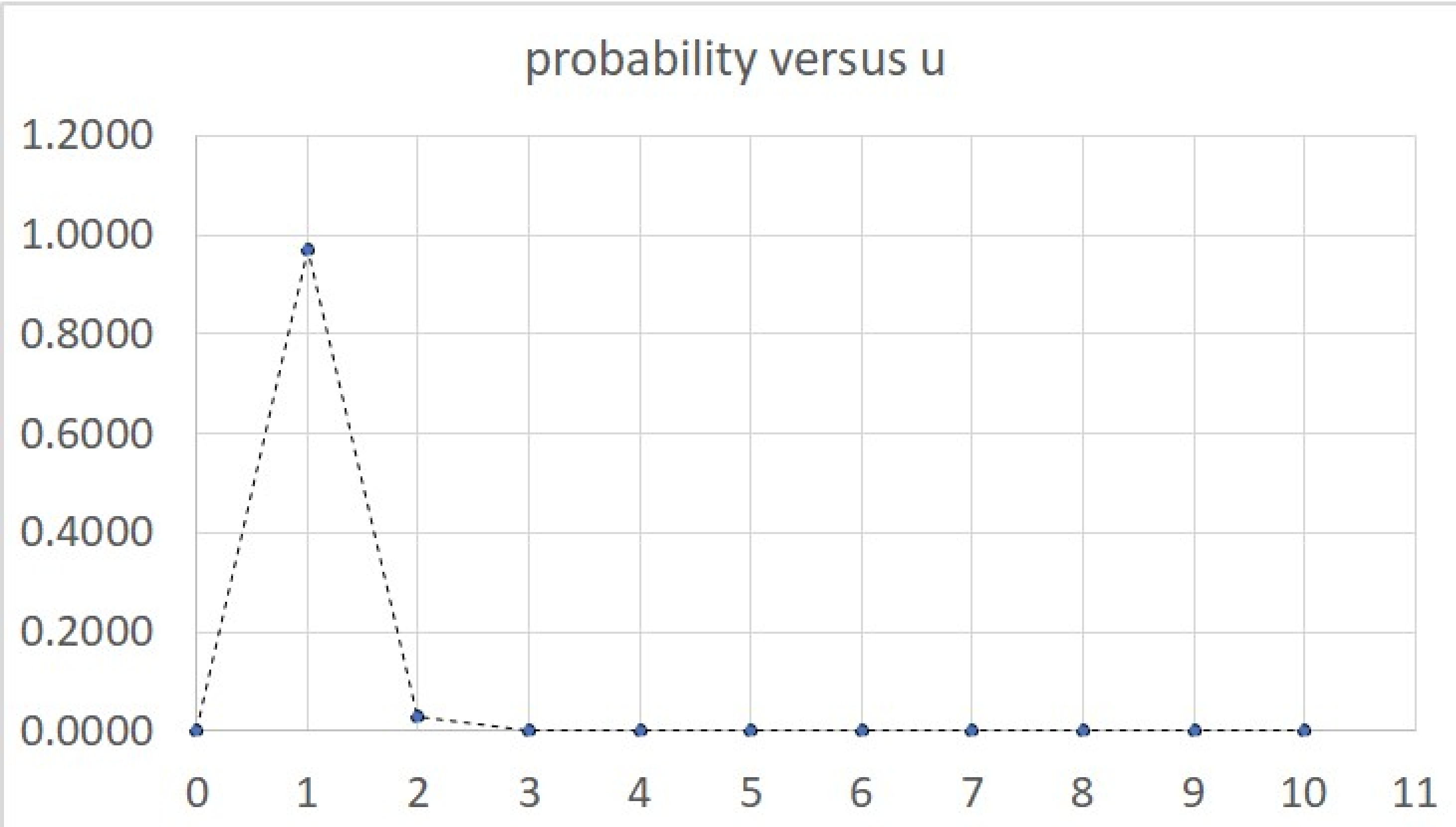


N	nB	u	fu=u/10	P(nb|u,N)	P(u)	P(nb|u,N)*P(u)	
10	5	0	0.00	0.0000	0.0909	0.0000	0.0000
10	5	1	0.10	0.0015	0.0909	0.0001	0.0016
10	5	2	0.20	0.0264	0.0909	0.0024	0.0291
10	5	3	0.30	0.1029	0.0909	0.0094	0.1133
10	5	4	0.40	0.2007	0.0909	0.0182	0.2208
10	5	5	0.50	0.2461	0.0909	0.0224	0.2708
10	5	6	0.60	0.2007	0.0909	0.0182	0.2208
10	5	7	0.70	0.1029	0.0909	0.0094	0.1133
10	5	8	0.80	0.0264	0.0909	0.0024	0.0291
10	5	9	0.90	0.0015	0.0909	0.0001	0.0016
10	5	10	1.00	0.0000	0.0909	0.0000	0.0000
						0.0826	



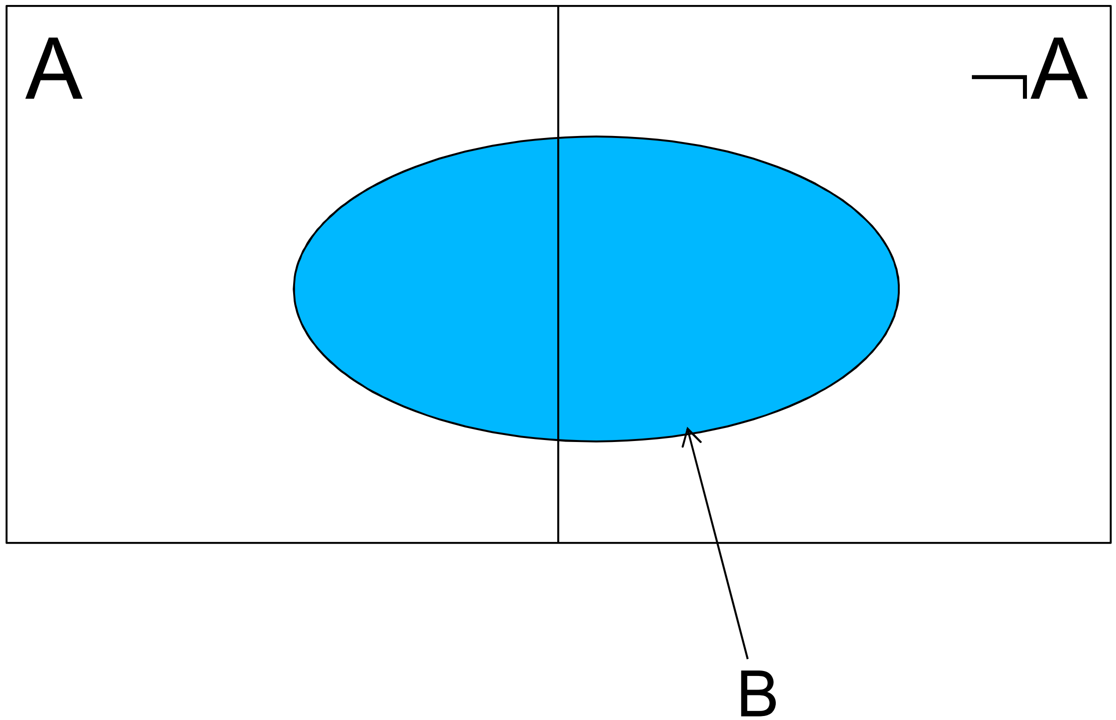
In 10 drawings, we have got 5 black balls.
Probability that we could have chosen bag#5 is around 0.27

In 100 drawings, we have got 3 black balls.
Probability that we could have chosen bag#1 is around 1



$$B = (A \cap B) + (\neg A \cap B)$$

$$P(B) = P(A \cap B) + P(\neg A \cap B)$$



$$\boxed{p(A|B) = \frac{p(B|A) \times p(A)}{p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)}}$$

Bayes' theorem – e.g.1

- Suppose that an individual's probability of having cancer, assigned according to the general prevalence of cancer, is 1%
- This is known as the "base rate" or prior (i.e. before being informed about the particular case at hand) probability of having cancer
- Writing *C* for the event "having cancer",
  - $P(C)=0.01$
- Suppose also that the probability of being 65 years old is 0.2%
  - $P(65)=0.002$
- Let us suppose next that cancer and age are related in the following way: the probability for someone diagnosed with cancer to be 65 is 0.5%
  - $P(65|C)=0.005$

Bayes' theorem – e.g.

- Calculate the probability of having cancer as a 65-year-old i.e. $P(C|65)$

$$P(C|65)= [P(65|C) * P(C)]/P(65)$$

$$=[0.005*0.01]/0.002=0.025=2.5\%$$

- Doctors find that people with lever disease almost invariably drunkards,

  $p(drunkard|\text{cancer}) = 0.9$

- The probability of an individual having *cancer* is currently rather low, about one in 100 000
- Assuming drinking alcohol is rather widespread, say $p(drunkard) = 0.5$, what is the probability that a drunkard will have cancer?
- $p(\text{cancer}|drunkard)$

$=p(drunkard|\text{cancer})p(\text{cancer})/p(drunkard)$

$=(9/10)\text{x}(1/100000)/(1/2)= 1.8 \times 10^{-5}$

- If the fraction of people drinking alcohol is small, $p(drunkard) = 0.001$, what is the probability that a *drunkard* will have cancer?

$p(drunkard|\text{cancer})p(\text{cancer})/p(drunkard)$

$=(9/10)\text{x}(1/100000)/(0.001)= 0.009$

- Two fair dice are rolled. Someone tells you that the sum of the two scores is 9. What is the posterior distribution of the dice scores?
- Score of die *a* is denoted s_a and score of die *b* is denoted s_b
- Total score, $t = s_a + s_b$
- A model of these three variables naturally takes the form $p(t, s_a, s_b) = p(t \mid (s_a, s_b)) \times p(s_a, s_b)$
- Die a and die b are independent i.e. $p(s_a, s_b) = p(s_a) x p(s_b)$

$$p(t, s_a, s_b) = p(t|s_a, s_b)\, p(s_a)\, p(s_b)$$

$$p(s_a, s_b) = p(s_a)p(s_b)$$

$$p(s_a)p(s_b):$$

	$s_a = 1$	$s_a = 2$	$s_a = 3$	$s_a = 4$	$s_a = 5$	$s_a = 6$
$s_b = 1$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
$s_b = 2$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
$s_b = 3$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
$s_b = 4$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
$s_b = 5$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
$s_b = 6$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$

$$p(s_a) = p(s_b) = 1/6$$

$$p(t = 9 | s_a, s_b):$$

	$s_a = 1$	$s_a = 2$	$s_a = 3$	$s_a = 4$	$s_a = 5$	$s_a = 6$
$s_b = 1$	0	0	0	0	0	0
$s_b = 2$	0	0	0	0	0	0
$s_b = 3$	0	0	0	0	0	1
$s_b = 4$	0	0	0	0	1	0
$s_b = 5$	0	0	0	1	0	0
$s_b = 6$	0	0	1	0	0	0

$$p(t = 9|s_a, s_b)p(s_a)p(s_b):$$

	$s_a = 1$	$s_a = 2$	$s_a = 3$	$s_a = 4$	$s_a = 5$	$s_a = 6$
$s_b = 1$	0	0	0	0	0	0
$s_b = 2$	0	0	0	0	0	0
$s_b = 3$	0	0	0	0	0	$1/36$
$s_b = 4$	0	0	0	0	$1/36$	0
$s_b = 5$	0	0	0	$1/36$	0	0
$s_b = 6$	0	0	$1/36$	0	0	0

$$p(t=9) = \sum_{s_a, s_b} p(t=9|s_a, s_b) p(s_a) p(s_b) = 1/9$$

$$p(s_a, s_b|t=9) = \frac{p(t=9|s_a, s_b) p(s_a) p(s_b)}{p(t=9)}$$

	$s_a = 1$	$s_a = 2$	$s_a = 3$	$s_a = 4$	$s_a = 5$	$s_a = 6$
$s_b = 1$	0	0	0	0	0	0
$s_b = 2$	0	0	0	0	0	0
$s_b = 3$	0	0	0	0	0	1/4
$s_b = 4$	0	0	0	0	1/4	0
$s_b = 5$	0	0	0	1/4	0	0
$s_b = 6$	0	0	1/4	0	0	0