So, compared to sampling with replacement (the binomial case), the VARIANCE of the number of Whites is

- SMALLER when sampling without replacement (Hypergeometric case)
- BIGGER when sampling according to the Polya scheme (beta-binomial case)

INDEPENDENCE

Two events A and B on the same sample space S (assume for now a "classical prob scenario") are said to be independent IF and ONLY IF

$$P(A \cap B) = P(A)P(B) \qquad \text{———} \qquad (*)$$

Two RVs $X(\cdot)$ and $Y(\cdot)$ defined on the same sample space are said to be independent if and only if for each (x, y), the events "$X = x$" and "$Y = y$" are independent events. (In practice we restrict attention to "possible values" x and y for $X(\cdot)$ and $Y(\cdot)$ respectively although the ~~df~~ definition still "makes sense" for any real x and y; if either is "not possible" then both sides of $(*)$ above become 0).

So if X and Y are indep, for each (x,y)

$$P(X=x, Y=y) = P(X=x)P(Y=y) \quad . \qquad (\dagger)$$

The LHS of $(\dagger)$ is called the joint prob. dist'n (or joint prob. mass function, or joint prob function). Thus RVs are indep if and only if their joint prob. dist'n is the product of their individual (or marginal) prob. dist'ns.

Examples: ① 2 6-sided dice (say, ~~red one~~ red and green) are rolled s.t. each of the 36 possible pairs of values is equally likely to be showing face-up. Let X = # showing on red, Y = # showing on green. Then for each $x, y = 1, 2, \ldots, 6$,

$$P(X=x) = P(\{(x,1),(x,2),(x,3),\ldots,(x,6)\}) = \frac{6}{36} = \frac{1}{6}$$

$$= P(Y=y) \text{ for } y = 1, 2, \ldots, 6.$$

and
$$P(X=x, Y=y) = P(\{(x,y)\}) = \frac{1}{36} = P(X=x)P(Y=y)$$

So X and Y are independent.

(2) An urn contains w white and b black balls, 2 are drawn out (assume $w \geq 1$, $b \geq 1$)

i) with replacement;
ii) without replacement;
iii) according to the Polya scheme

in such a way that each possible seq. is equally likely.

Let X = # white balls on 1ST draw, Y = # wh. balls on 2ND draw.

Write the joint prob distn and both marginal distns in each case. Are X and Y independent?

i) WITH REPL. Total # possible sequences is $(w+b)^2$

sequences with BB is b^2

.. .. .. BW is bw

.. .. .. WB is wb

.. .. .. WW is w^2

$$P(X=0, Y=0) = P(BB) = \frac{b^2}{(w+b)^2}$$

$$P(X=0, Y=1) = P(BW) = \frac{bw}{(w+b)^2}$$

$$P(X=1, Y=0) = P(WB) = \frac{wb}{(w+b)^2}$$

$$P(X=1, Y=1) = P(WW) = \frac{w^2}{(w+b)^2}$$

	$Y=0$	$Y=1$	Total
$X=0$	$\left(\frac{b}{w+b}\right)^2$	$\frac{bw}{(w+b)^2}$	$\frac{b^2+bw}{(w+b)^2} = \frac{b\cancel{(b+w)}}{(w+b)^{\cancel{2}}} = \frac{b}{w+b}$
$X=1$	$\frac{wb}{(w+b)^2}$	$\frac{w^2}{(w+b)^2}$	$\frac{w(w+b)}{(w+b)^2} = \frac{w}{w+b}$
total	$\frac{b}{w+b}$	$\frac{w}{w+b}$	1

Note $P(X=0) = P(Y=0) = \frac{b}{w+b}$

$P(X=1) = P(Y=1) = \frac{w}{w+b}$

AND Note

$$P(X=0, Y=0) = P(X=0)P(Y=0)$$
$$P(X=0, Y=1) = P(X=0)P(Y=1)$$
$$P(X=1, Y=0) = P(X=1)P(Y=0)$$
$$P(X=1, Y=1) = P(X=1)P(Y=1)$$

So X and Y are INDEPENDENT.

ii) <u>WITHOUT REPL.</u> Total # sequences: $(w+b)(w+b-1)$

seq with BB : $b(b-1)$

" " BW : bw

" " WB : wb

" " WW : $w(w-1)$

	$Y=0$	$Y=1$	Total
$X=0$	$\frac{b(b-1)}{(w+b)(w+b-1)}$	$\frac{bw}{(w+b)(w+b-1)}$	$\frac{b^2-b+bw}{(w+b)(w+b-1)} = \frac{b(w+b-1)}{(w+b)(w+b-1)} = \frac{b}{b+w}$
$X=1$	$\frac{wb}{(w+b)(w+b+1)}$	$\frac{w(w-1)}{(w+b)(w+b-1)}$	$\frac{w}{w+b}$
Total	$\frac{b}{b+w}$	$\frac{w}{w+b}$	1

So $P(X=0,Y=0) = \frac{b(b-1)}{(w+b)(w+b-1)} < \frac{b^2}{(w+b)^2} = P(X=0)P(Y=0)$

Check

Note that $\frac{b-1}{w+b-1} - \frac{b}{w+b} = \frac{(b-1)(w+b) - b(w+b-1)}{(\ldots\ldots)}$

$= \cancel{bw} - w + b^2 - b - [\cancel{bw} + b^2 - b]$

$= \frac{-w}{(\ldots\ldots)} < 0$

$$P(X=0,Y=1) = \frac{bw}{(w+b)(w+b-1)} > \frac{bw}{(w+b)^2} = P(X=0)P(Y=1)$$

Similarly $P(X=1,Y=0) > P(X=0)P(Y=1)$

and finally

$$P(X=1,Y=1) = \frac{w(w-1)}{(w+b)(w+b-1)} < \frac{w^2}{(w+b)^2} = P(X=1)P(Y=1)$$

So X and Y are NOT INDEPENDENT. More precisely,

- its more likely to have 2 different colours (compared to independence)
- its less likely to have 2 ~~a~~ colours the same (" " ")

<u>Beta-binomial</u>: [exercise for elsewhere]

<u>Variance of a SUM of RVs</u>; <u>Covariance</u>