Unit #8: Introduction to Bayesian Inference

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, 1)$. How would you estimate μ? Intuition says $\bar{X}$.

1. How "good" is this estimator? What does "good" even mean?

2. Can we do better?

How would you go about estimating parameters with a less obvious interpretation, such as the α from the $\Gamma(\alpha, \beta)$ distribution?

Frequentist answer: an estimator is "good" if it has certain properties—consistency, unbiasedness, efficiency—over repeated samples.

Bayesian answer: an estimator is "good" if it is derived from the posterior distribution of the parameter given the data.

$E(\bar{X}) = \mu$

$\bar{X} \xrightarrow{P} \mu$

The Big Picture

$\mu \sim U(5, 6)$

$X_1, \ldots, X_n \sim ?$

$\mu \mid \underline{x} \sim ?$

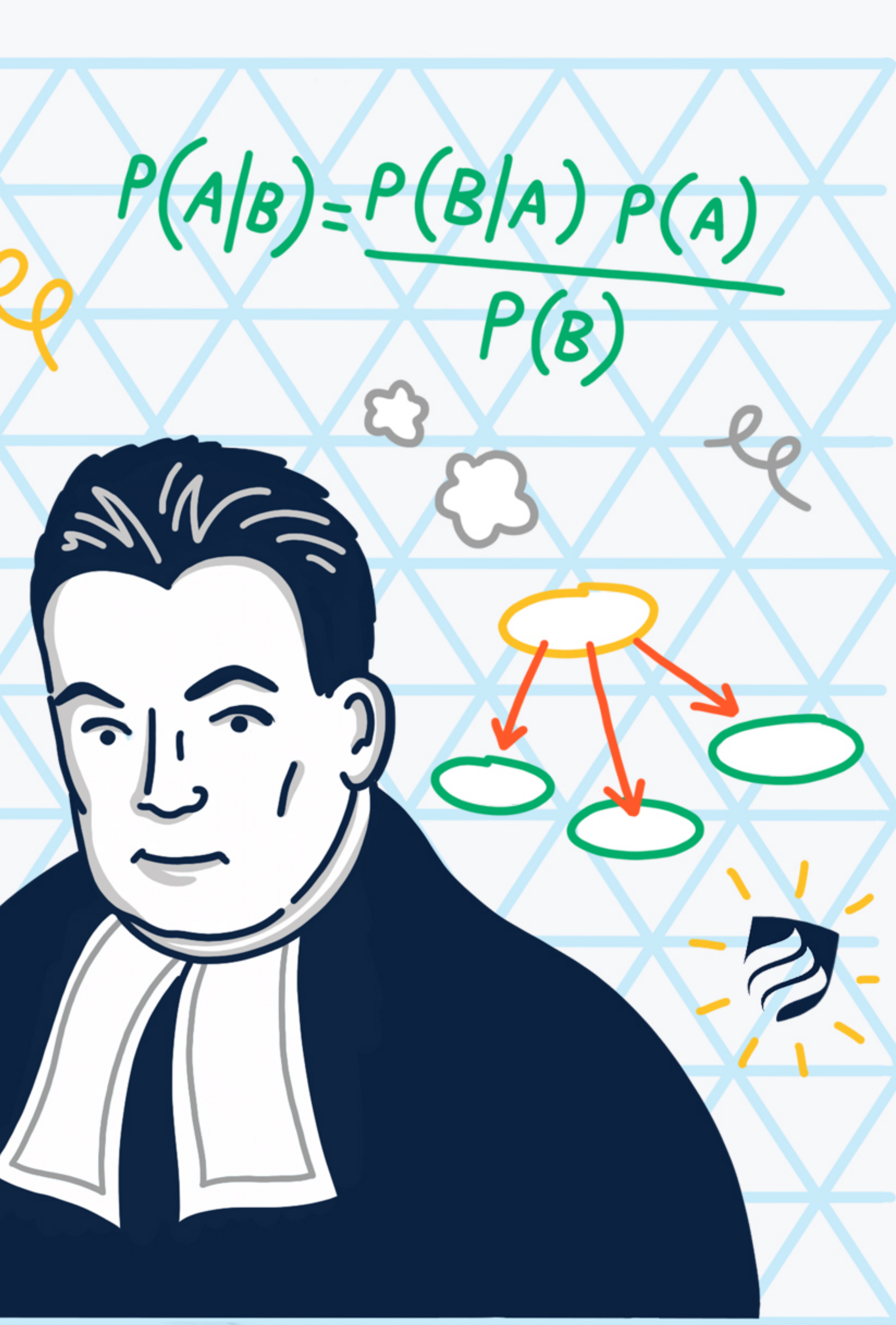
Broadly speaking, there are two different paradigms for finding estimators for parameters. The different paradigms result from different interpretations of probability.

Frequentist:

- Parameters are unknown and fixed.
- Estimators for parameters are derived from sample data.
- We assess the *frequentist properties* of the estimator: that is, is the estimator unbiased? Consistent? Efficient?
- These properties have to do w/*frequency,* i.e., repeated sampling.
  - E.g., if I chose a sample of size *n* over and over from this population, would my estimator be correct, on average?

Bayesian:

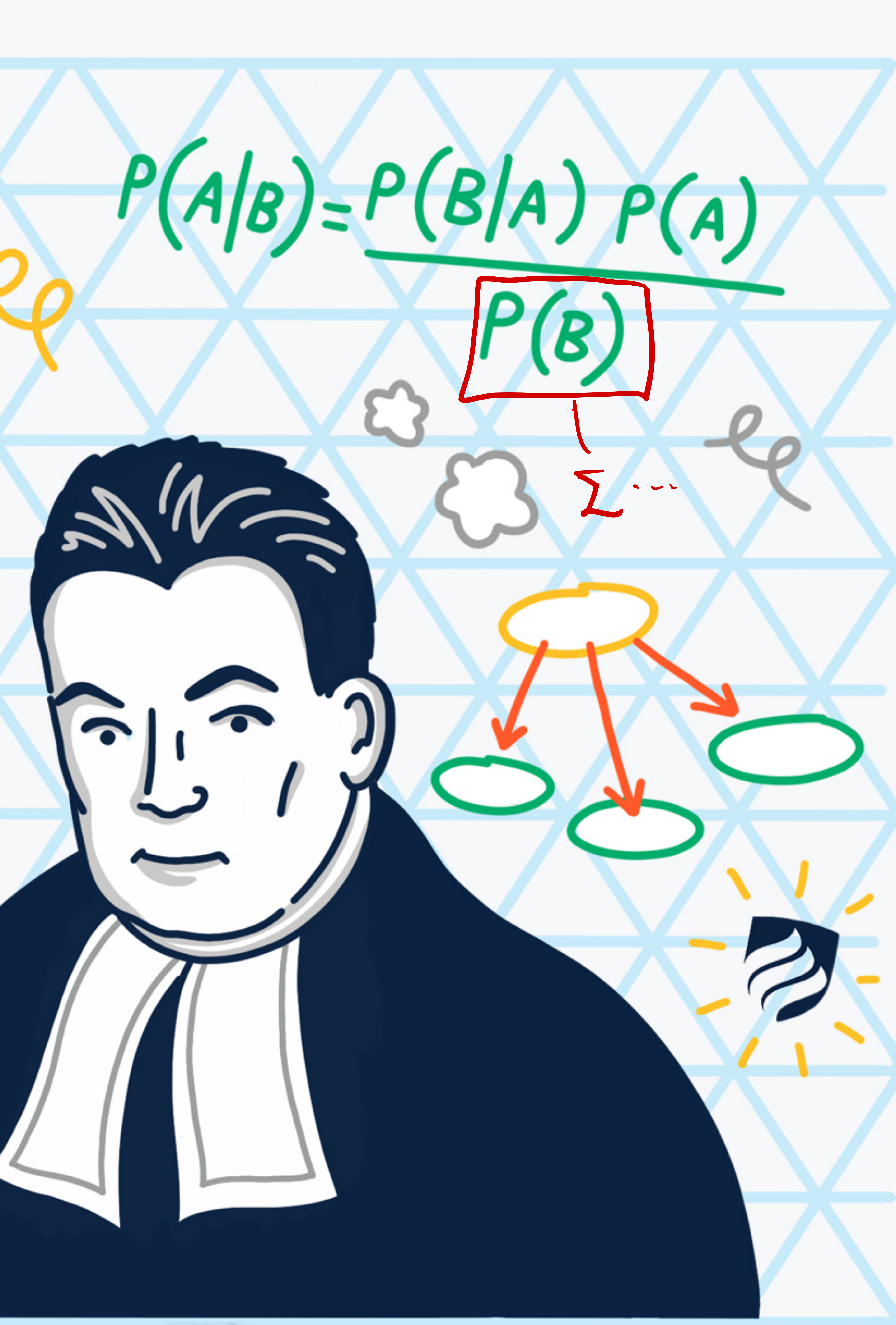
- Parameters are unknown and fixed, but we *model* them as random variables.
  - This model is usually thought of as representing our *degree of belief* in the value of the parameter.
  - We might think (before sampling): I don't know the (fixed) mean height of all CU students, but I believe that it's definitely between 5 and 6 ft. (Prior Belief).
- We use Bayes' theorem to combine our prior belief about a parameter with evidence from a sample.
- The result is a new probability distribution over the parameter. It represents our belief in the value of the parameter after accounting for evidence (the sample). The mean or mode might serve as an est.



From: https://bit.ly/3gTjlBg

In the frequentist approach to estimation (what we've looked at so far), parameters, θ, were thought to be unknown and *fixed* (i.e., not random variables).

In the Bayesian approach, parameters are considered to be a quantity whose variation can be described by a (subjective) probability distribution (called a *prior distribution*). A sample is then taken from the population and the prior is updated—using Bayes' theorem—to obtain a *posterior distribution*.



From: https://bit.ly/3gTjlBg

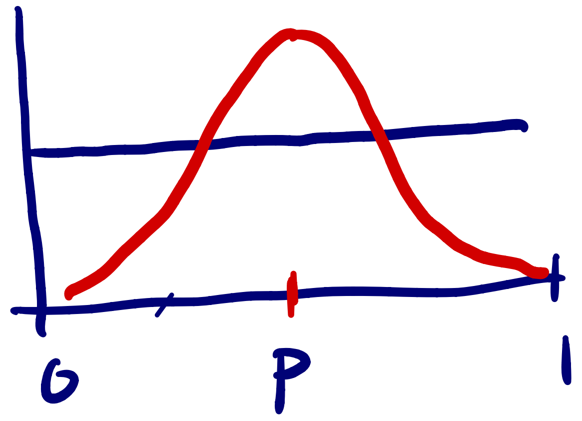
Denote the *prior distribution* by $\pi(\theta)$. Then the *posterior distribution*, the conditional distribution of θ given the sample $\mathbf{x}$, is: $\left(\underline{x} = (x_1, \dots, x_n)\right)$

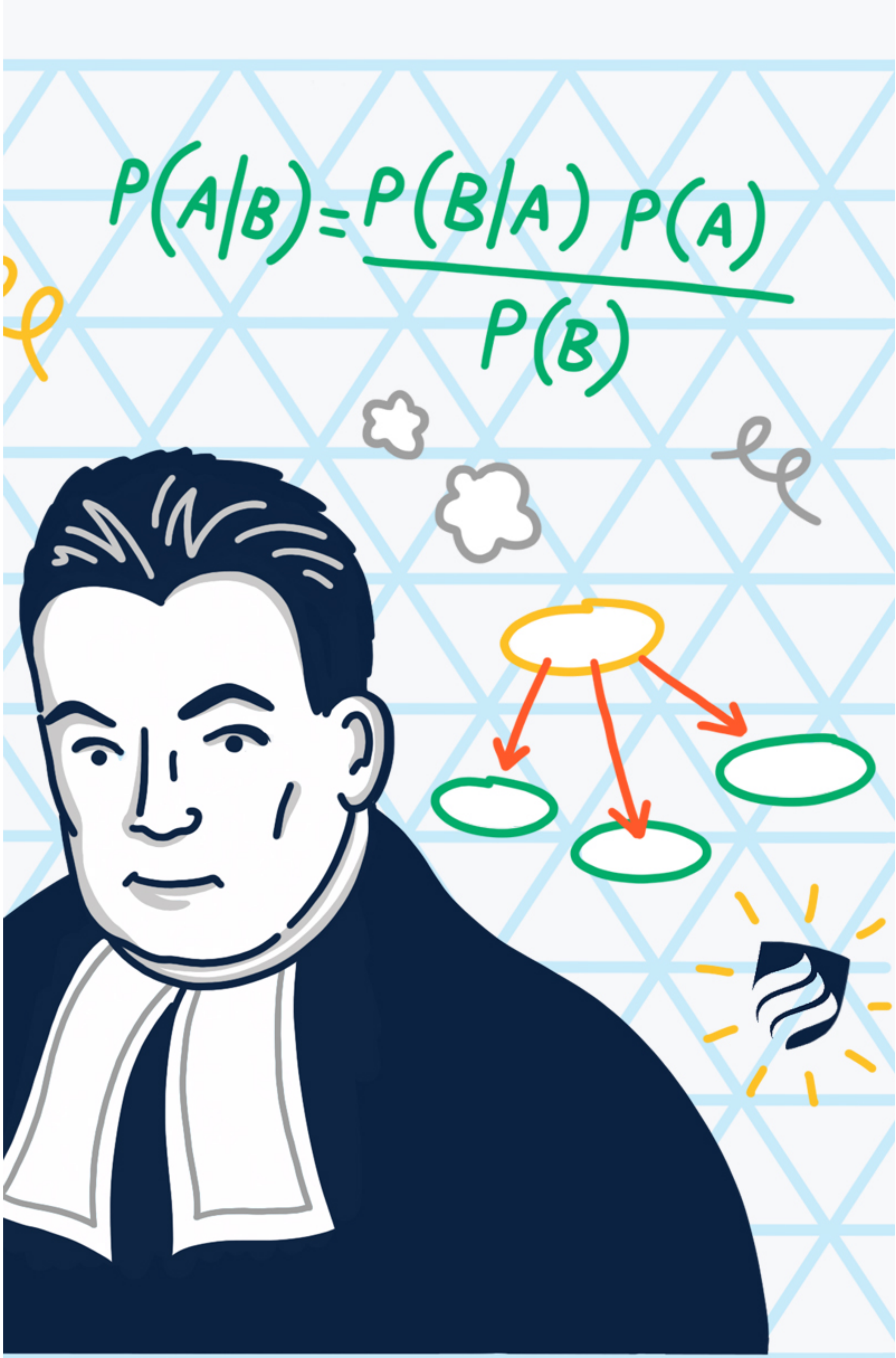
likelihood function (doesn't integrate to 1 of θ)

$$\underbrace{\pi(\theta \mid \underline{x})}_{\int \pi(\theta|\underline{x})\,d\theta = 1} = \frac{\boxed{f(\underline{x} \mid \theta)}\,\pi(\theta)}{\boxed{\int_{-\infty}^{\infty} f(\underline{x} \mid \theta)\,\pi(\theta)\,d\theta}}$$

constant WRT θ

Note: the posterior distribution is used to make statements about θ, which is still considered a random quantity. For example, the mean or mode of the posterior can be used as a point estimate of θ.





From: https://bit.ly/3gTjlBg

Example: Binomial Bayes Estimation. Let $X_1, \ldots, X_n \overset{iid}{\sim} Binomial(1,p)$. Let the prior distribution on p be a beta distribution. Find the posterior distribution of $p \mid \mathbf{x}$, i.e., the posterior distribution of the parameter p given that we observed the data $\mathbf{x}$.

- Prior distribution: $\pi(p) = \overbrace{\boxed{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}}^{\text{const.}} p^{\alpha-1}(1-p)^{\beta-1}$, for $0 < p < 1$
- Likelihood: marginal pmf of $Bin(1,p)$: $f(x_i \mid p) = p^{x_i}(1-p)^{1-x_i}$ $\quad x_i = 0, 1$

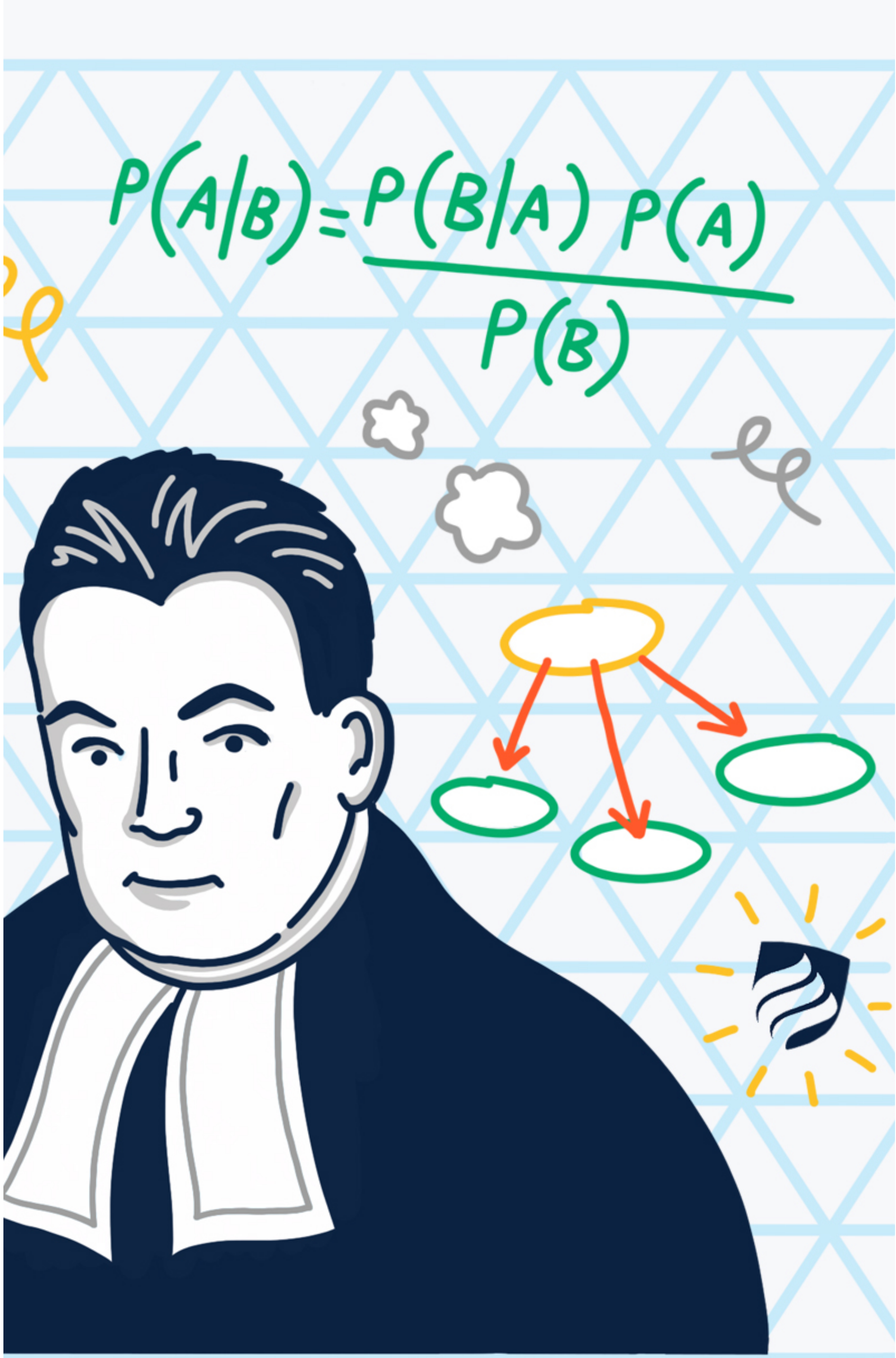
  $\hookrightarrow f(\underline{x} \mid p) \overset{i}{=} \prod_{i=1}^{n} f(x_i \mid p) = p^{\sum x_i}(1-p)^{n-\sum x_i} = p^{n\bar{x}}(1-p)^{n-n\bar{x}}$
- Posterior:

$$\pi(p \mid \underline{x}) \propto f(\underline{x} \mid p)\pi(p) = p^{n\bar{x}}(1-p)^{n-n\bar{x}}\, p^{\alpha-1}(1-p)^{\beta-1}$$

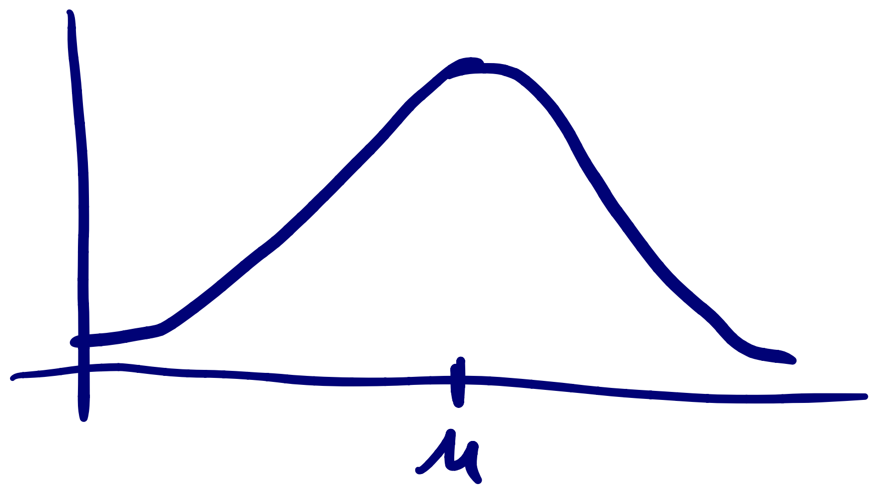
proportional to

$$= p^{n\bar{x}+\alpha-1}(1-p)^{n-n\bar{x}+\beta-1}$$

$$\Rightarrow p \mid \underline{x} \sim Beta(n\bar{x}+\alpha,\ n-n\bar{x}+\beta)$$



From: https://bit.ly/3gTjlBg



$E(X) = \mu$

Example: Binomial Bayes Estimation continued.

Let $X_1, \ldots, X_n \overset{iid}{\sim} Binomial(1, p)$. Let the prior distribution on p be a beta distribution. The posterior is given by:

$$p \mid \mathbf{x} \sim Beta\left(\underbrace{n\bar{x} + \alpha}_{\alpha_1}, \underbrace{n - n\bar{x} + \beta}_{\beta_1}\right)$$

$E(p \mid \underline{x}) = \dfrac{\alpha_1}{\alpha_1 + \beta_1}$

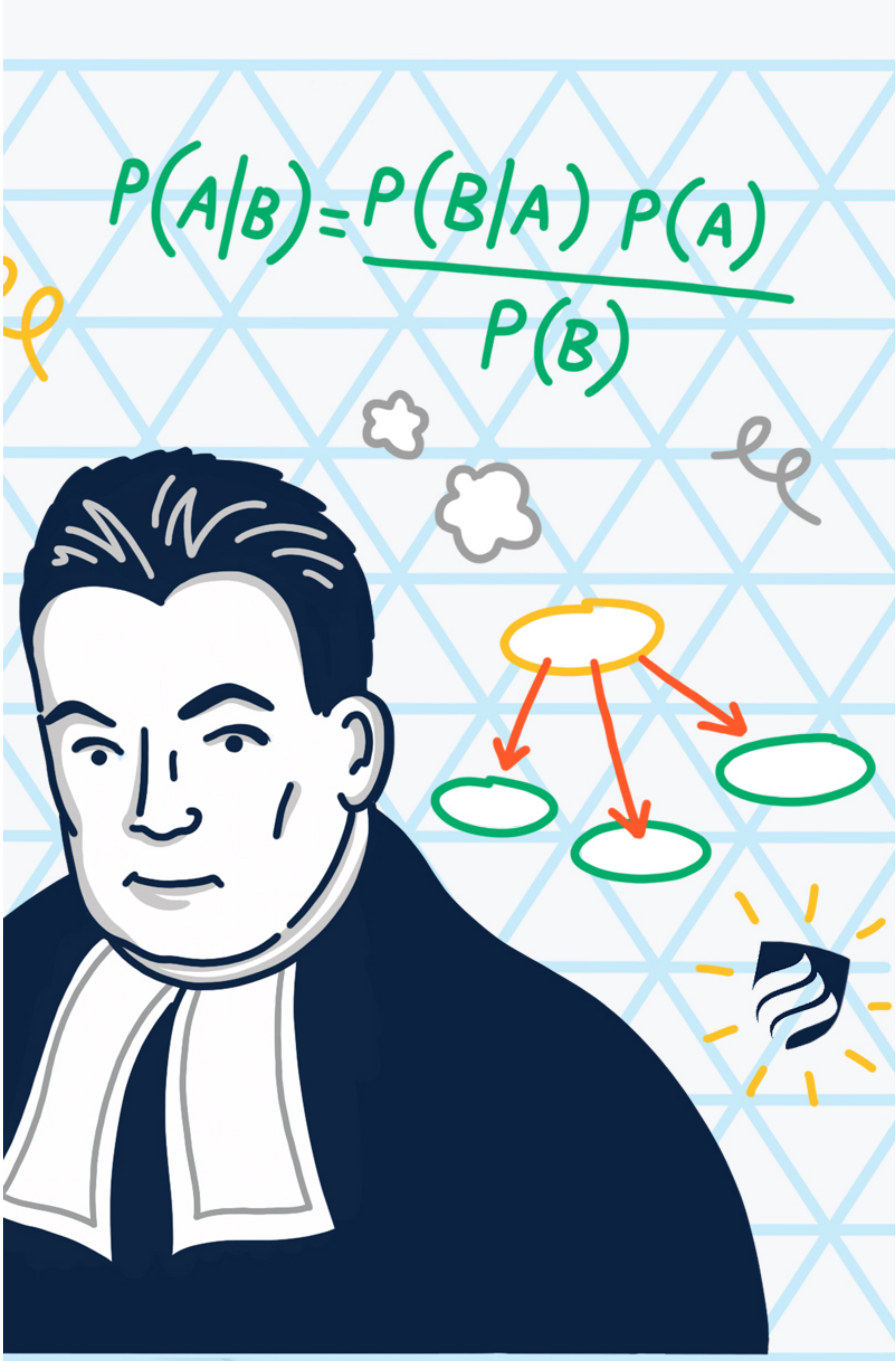
Estimator of p?

$$\hat{\theta}_1 = E(p \mid \mathbf{x}) = \frac{n\bar{x} + \alpha}{n + \alpha + \beta}$$

$\hat{\theta}_1 = \hat{p}_1$

$\left(\hat{p}_{ML} = \bar{X}\right)$

$\alpha = 1 = \beta \implies \hat{p}_0 = \frac{1}{2}$



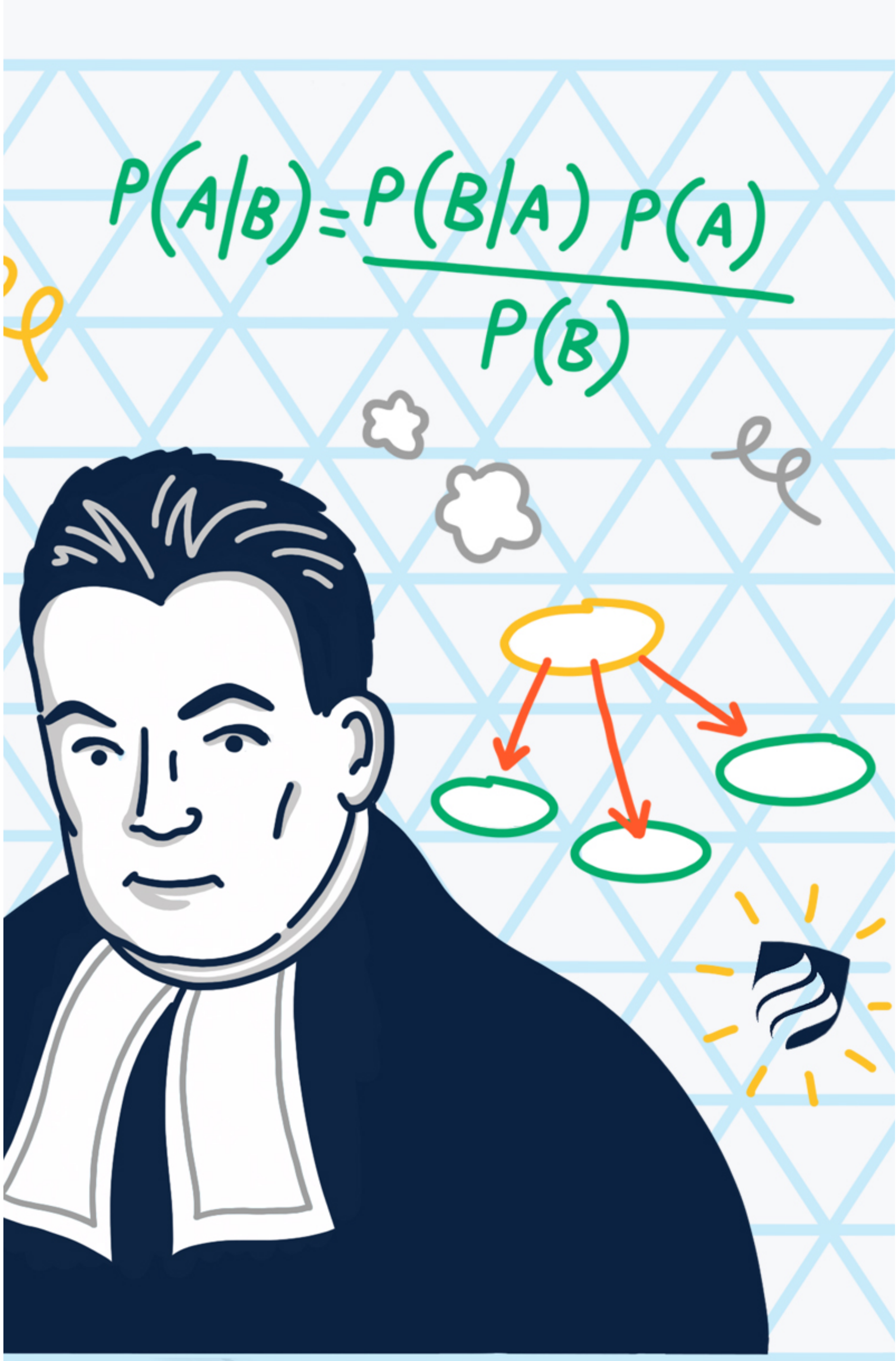
From: https://bit.ly/3gTjlBg

Example: Binomial Bayes Estimation continued.

Let $\hat{p}_0 = E(p) = \dfrac{\alpha}{\alpha + \beta}$ be the prior mean, and $\underbrace{\hat{p} = \bar{x}}_{\text{MLE}}$. Then the posterior mean can also be written as:

$$E(p \mid \mathbf{x}) = \frac{n\bar{x} + \alpha}{n + \alpha + \beta} = (1 - w_n)\hat{p}_0 + w_n\hat{p},$$

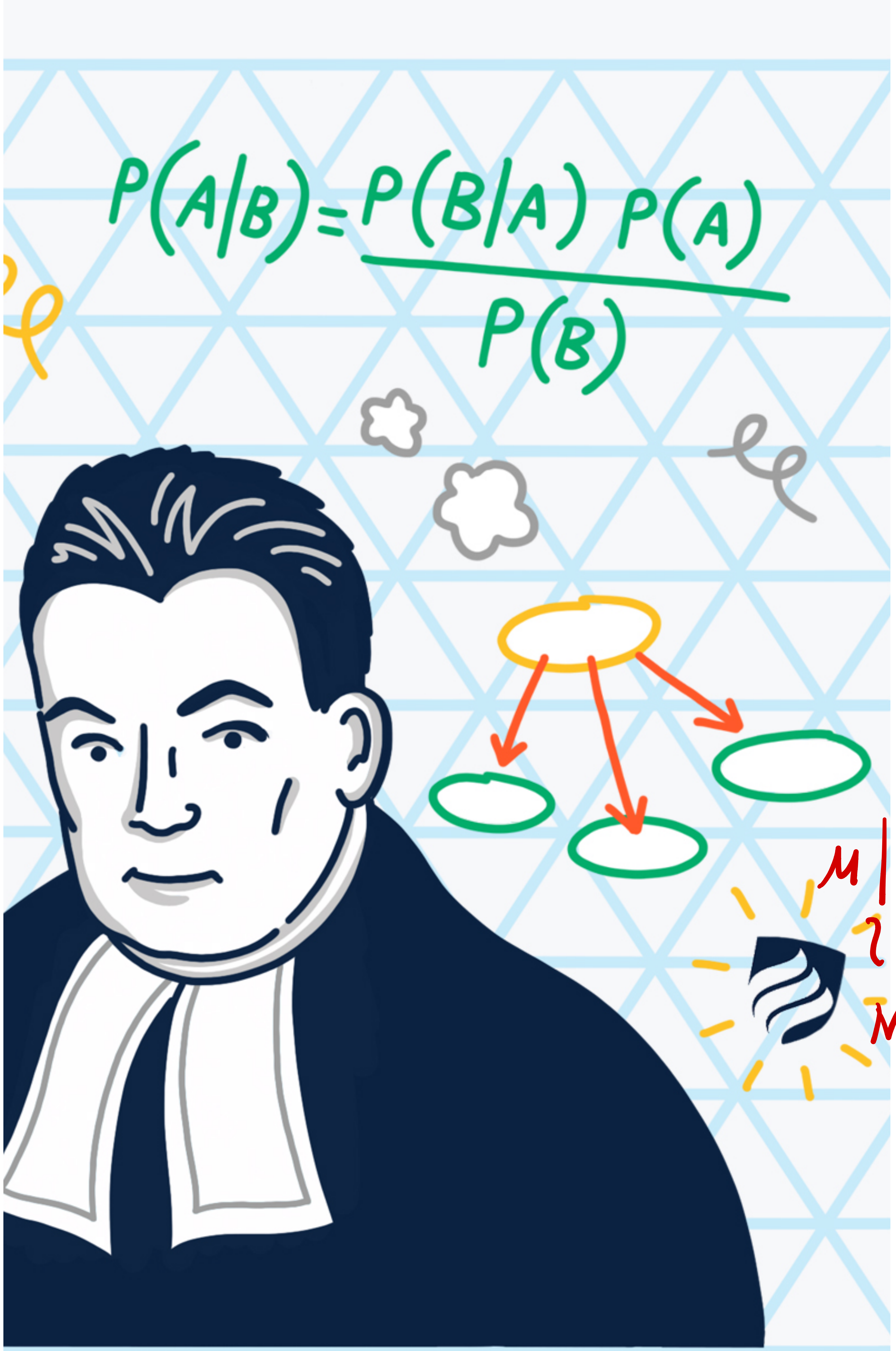
where $w_n = \dfrac{n}{n + \alpha + \beta}$.



From: https://bit.ly/3gTjlBg

Definition: A class of prior distributions, C, is a *conjugate family* for a likelihood, $L(\theta \mid \mathbf{x})$, if the posterior distribution is also in class C.

Example: In the previous example, we saw that the beta family of distributions (prior) is a conjugate for the binomial family (likelihood).



From: https://bit.ly/3gTjIBg

Example: Conjugate Prior for Normal Distribution. Let $X \sim N(\mu, \underbrace{\sigma^2}_{\text{Known}})$ and $\mu \sim N(\mu_0, \sigma_0^2)$. Find the posterior distribution of $\mu \mid \mathbf{x}$.

single observation!

$$\pi(\mu \mid \underline{x}) \propto \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\exp\left\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2 - \frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2} + \frac{(\mu-\mu_0)^2}{\sigma_0^2}\right)\right\}$$

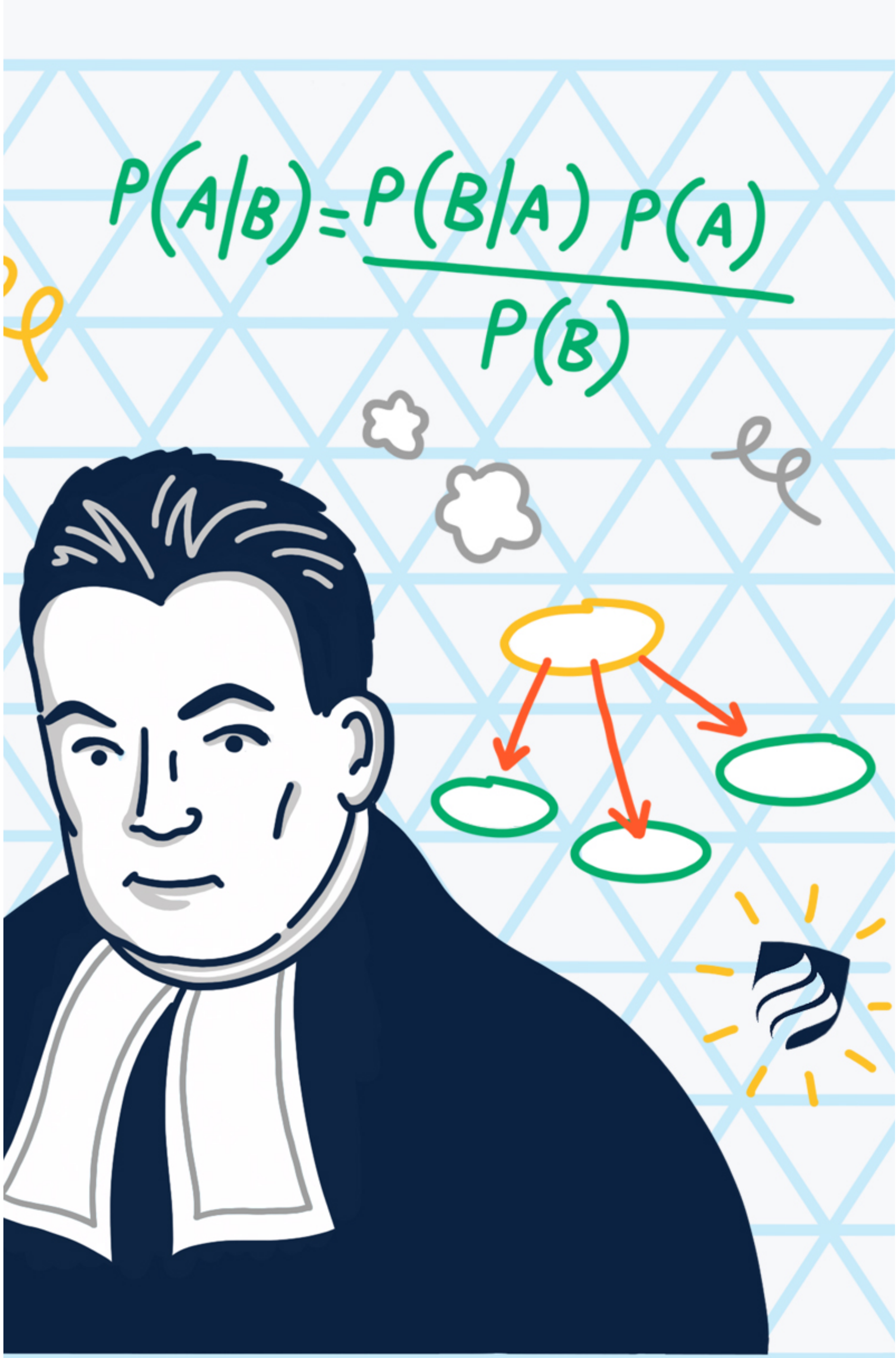
$$\propto \exp\left\{-\frac{1}{2}\left(\frac{x^2-2\mu x-\mu_0^2}{\sigma^2} + \frac{\mu^2-2\mu\mu_0+\mu_0^2}{\sigma_0^2}\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\mu^2 - 2\left(\frac{x}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)\mu + c\right]\right\}$$

complete square

$$\propto \exp\left\{-\frac{1}{2}\underbrace{\boxed{\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_0^2}\right)^{-1}}}_{\sigma_1^2}\left[\mu - \left(\underbrace{\boxed{\frac{x/\sigma^2 + \mu_0/\sigma_0^2}{1/\sigma^2 + 1/\sigma_0^2}}}_{\mu_1}\right)\right]^2\right\}$$

$\mu \mid x \sim N(\mu_1, \sigma_1^2)$



From: https://bit.ly/3gTjlBg

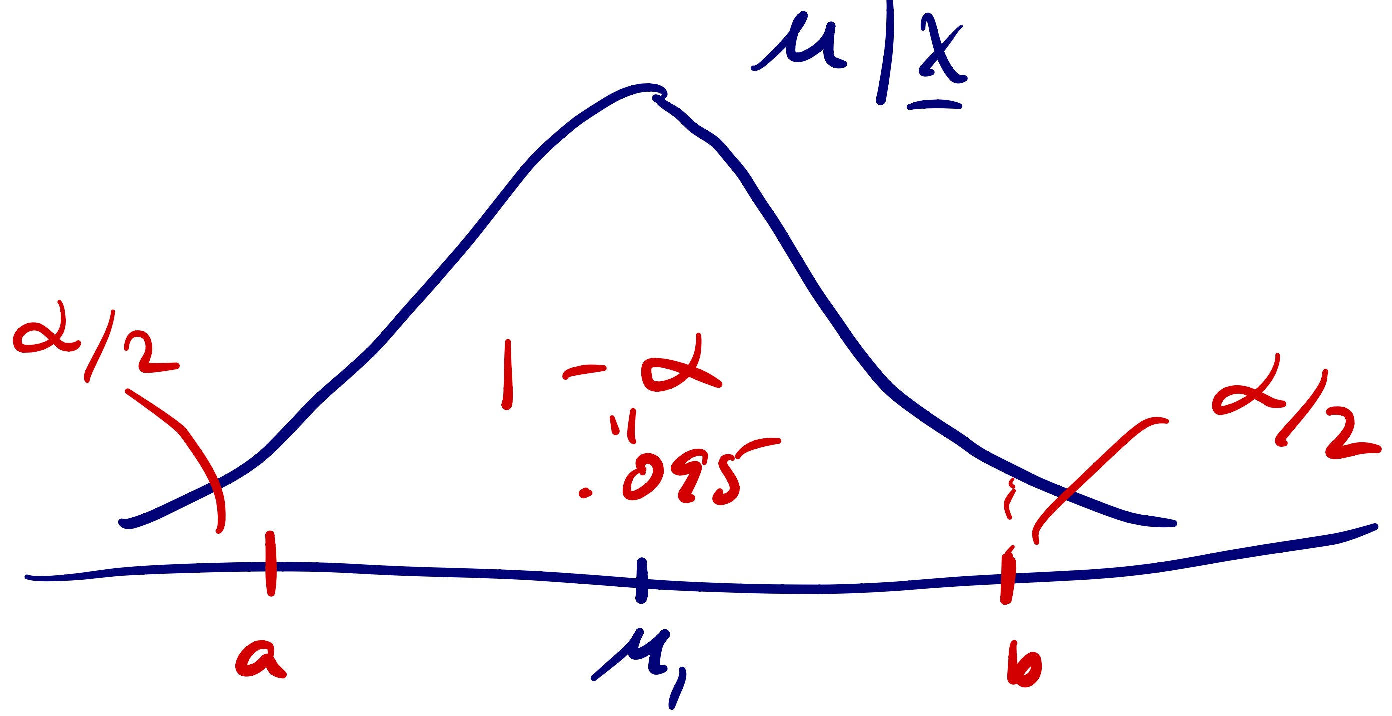
Example: Conjugate Prior for Normal Distribution. Let $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ and $\mu \sim N(\mu_0, \sigma_0^2)$. Find the ~~posterior~~ distribution of $\mu \mid \mathbf{x}$. $i = 1, \ldots, n$

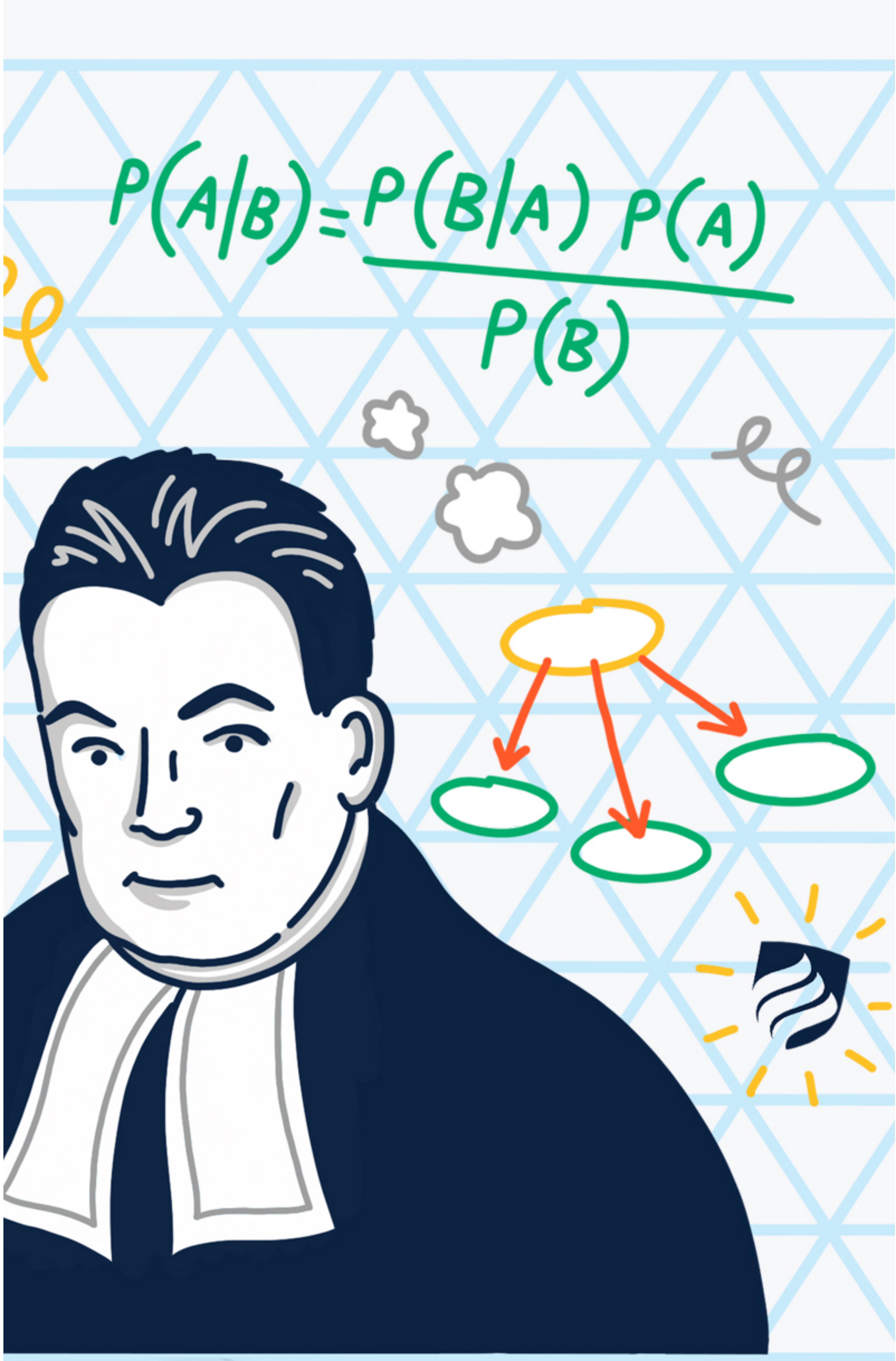
Multiple observations!

$$\mu \mid \underline{x} \sim N\left(\mu_1, \sigma_1^2\right)$$

$$\mu_1 = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \quad \sigma_1^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$

95% Credible interval:

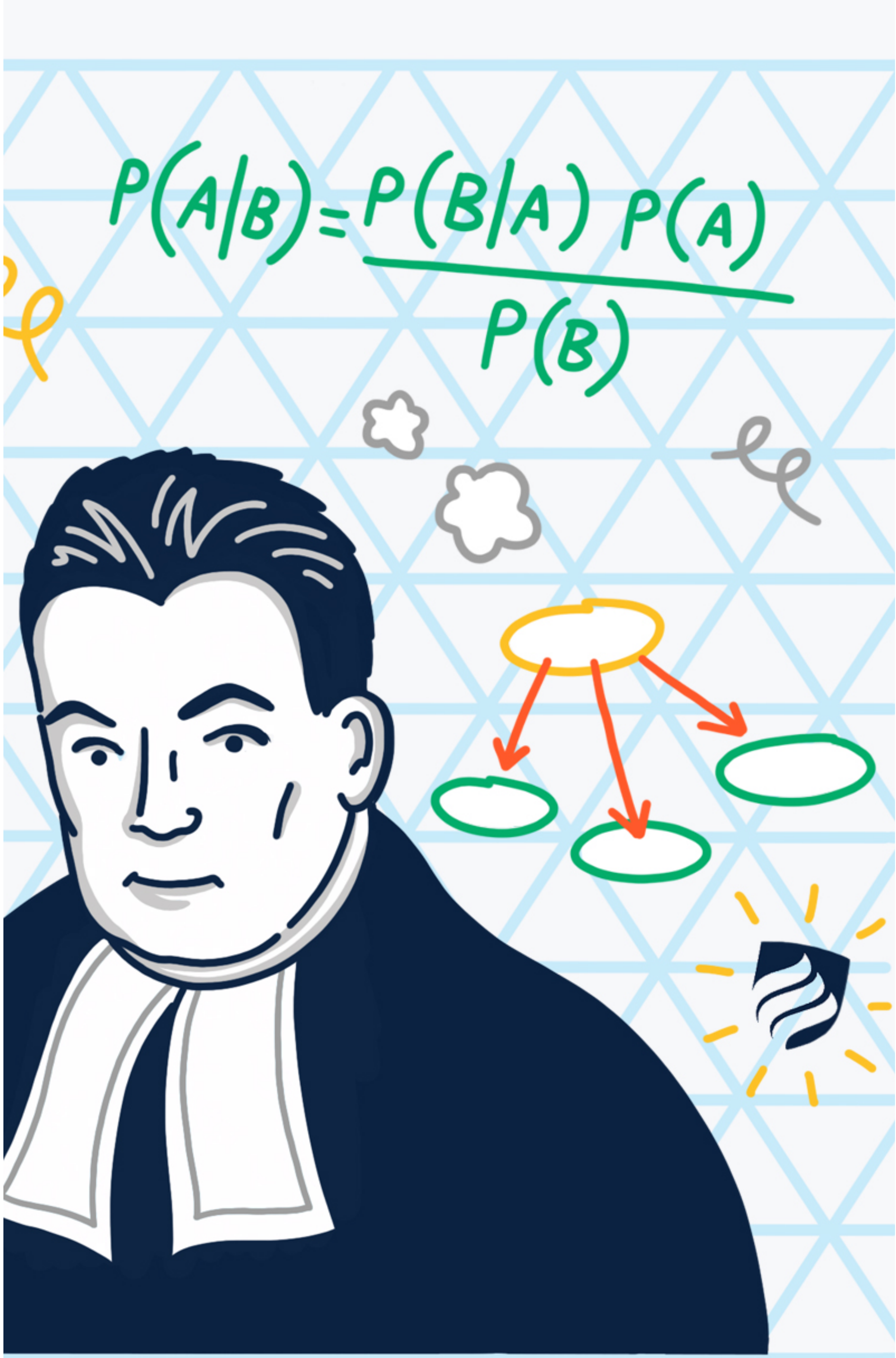




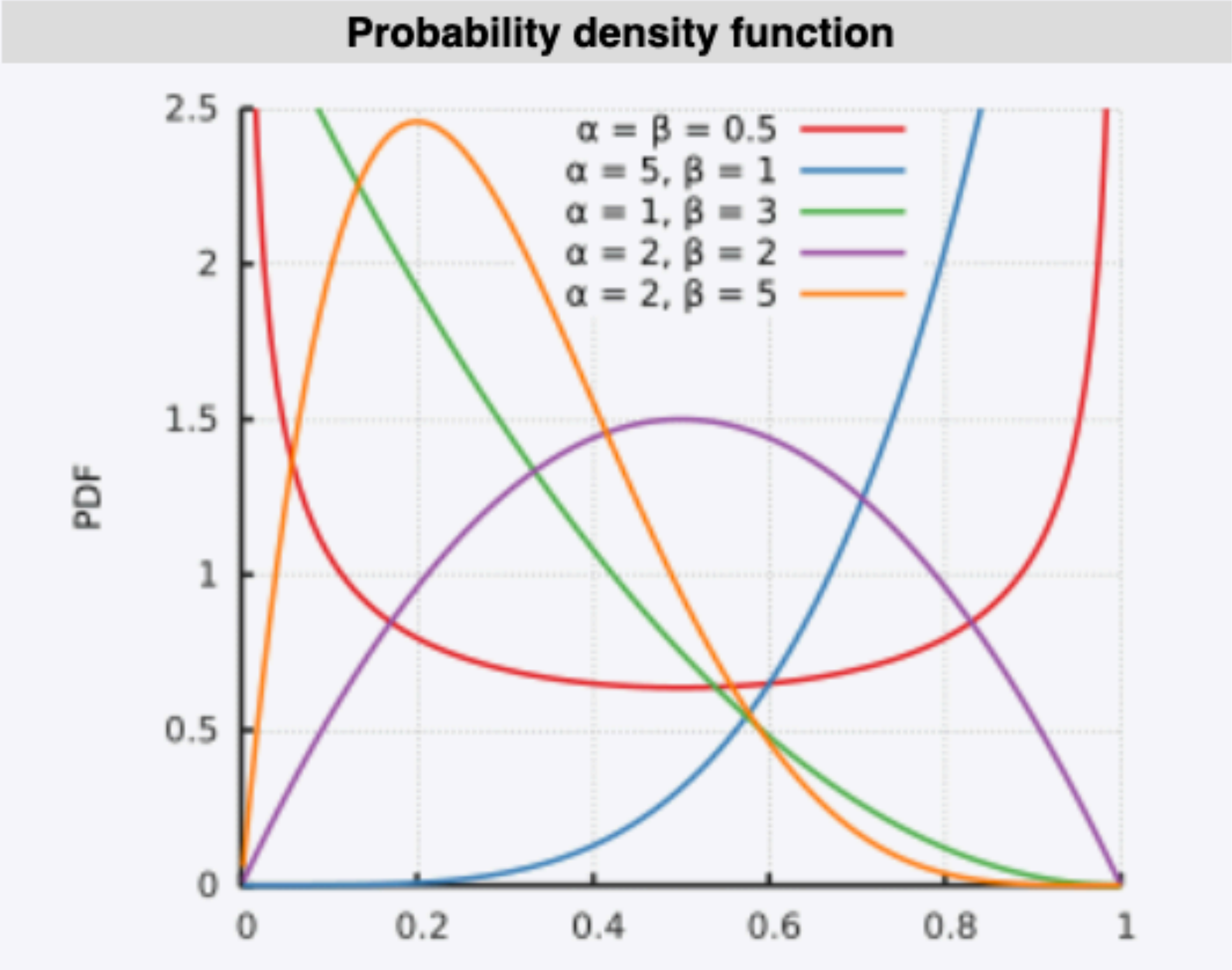
From: https://bit.ly/3gTjlBg

Example: Suppose researchers would like to know the probability, p, that an individual in a given population has a genetic marker that predisposes them for disease D (this genetic marker is such that an individual has it or doesn't; assume that our test for the marker is completely accurate). Researchers collected data in the following way: they tested people for this genetic marker until they found on person who had it. They stopped data collection at that point. Here's the data: $\mathbf{x} = (0,0,0,0,1)$. Suppose your prior beliefs are best represented by a beta$(\underset{\alpha}{2},\underset{\beta}{2})$ distribution. What is the posterior distribution for $p \mid x$?

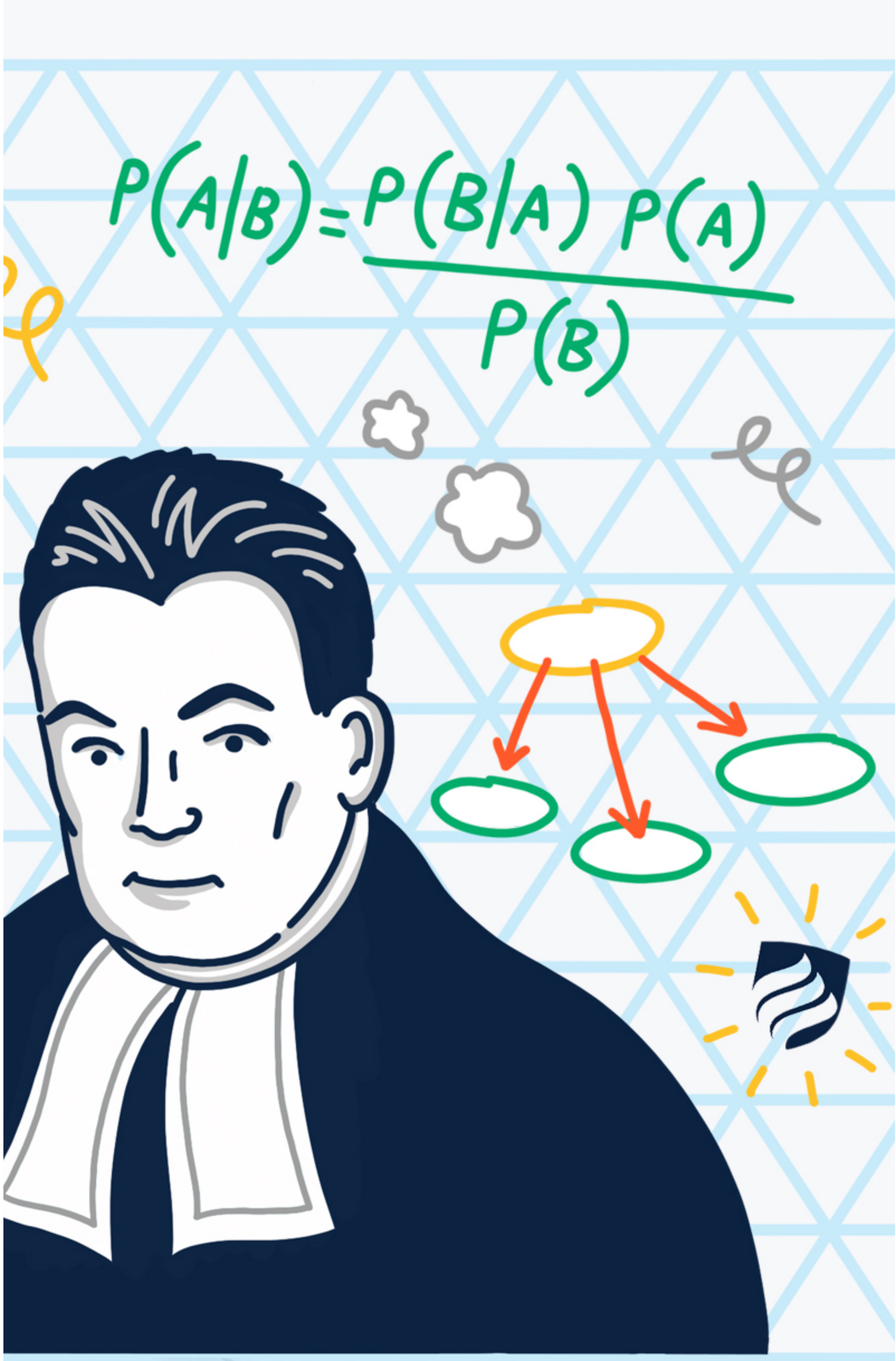
$p \sim \text{Beta}(2, 2)$, $E(p) = \frac{\alpha}{\alpha+\beta}$



From: https://bit.ly/3gTjlBg



Prior mean? $E(p) = \frac{1}{2}$



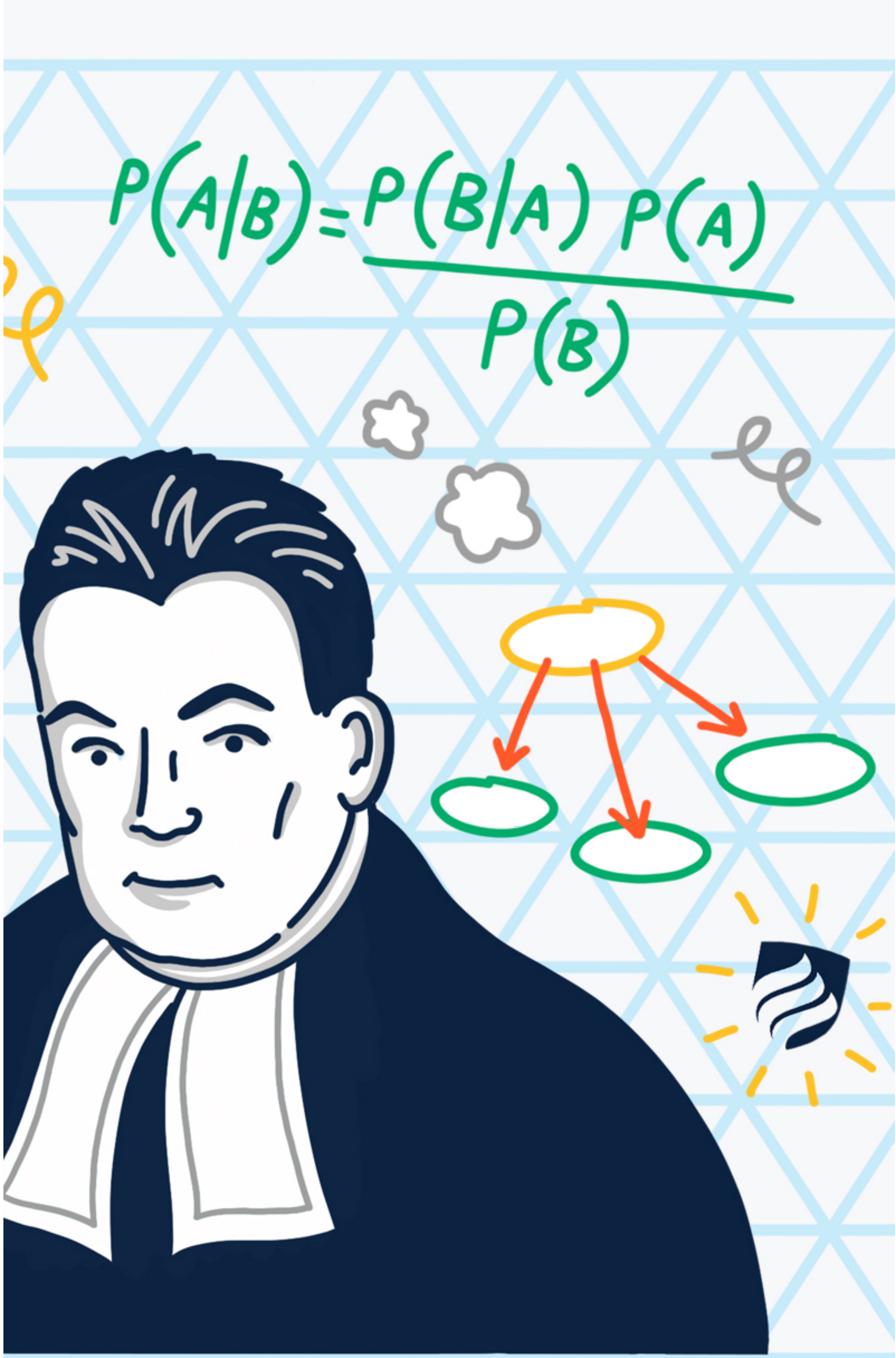
From: https://bit.ly/3gTjlBg

Example: Suppose researchers would like to know the probability, p, that an individual in a given population has a genetic marker that predisposes them for disease D (this genetic marker is such that an individual has it or doesn't; assume that our test for the marker is completely accurate). Researchers collected data in the following way: they tested people for this genetic marker until they found on person who had it. They stopped data collection at that point. Here's the data: $\mathbf{x} = (0,0,0,0,1)$. Suppose your prior beliefs are best represented by a beta$(2,2)$ distribution. What is the posterior distribution for $p \mid x$?

Geometric distribution: The probability distribution of the number of failures, Y, before the first success, supported on the set $\{0,1,2,...\}$

$$Y \sim G(p)$$

$$f(y \mid p) = (1-p)^y p \quad , \quad y \in \{0, 1, 2, \ldots\}$$



From: https://bit.ly/3gTjlBg

Example: Suppose researchers would like to know the probability, p, that an individual in a given population has a genetic marker that predisposes them for disease D (this genetic marker is such that an individual has it or doesn't; assume that our test for the marker is completely accurate). Researchers collected data in the following way: they tested people for this genetic marker until they found on person who had it. They stopped data collection at that point. Here's the data: $\mathbf{x} = (0,0,0,0,1)$. Suppose your prior beliefs are best represented by a beta$(2,2)$ distribution. What is the posterior distribution for $p \mid x$?

Prior: $\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} = \frac{\Gamma(4)}{\Gamma(2)\Gamma(2)} p(1-p)$

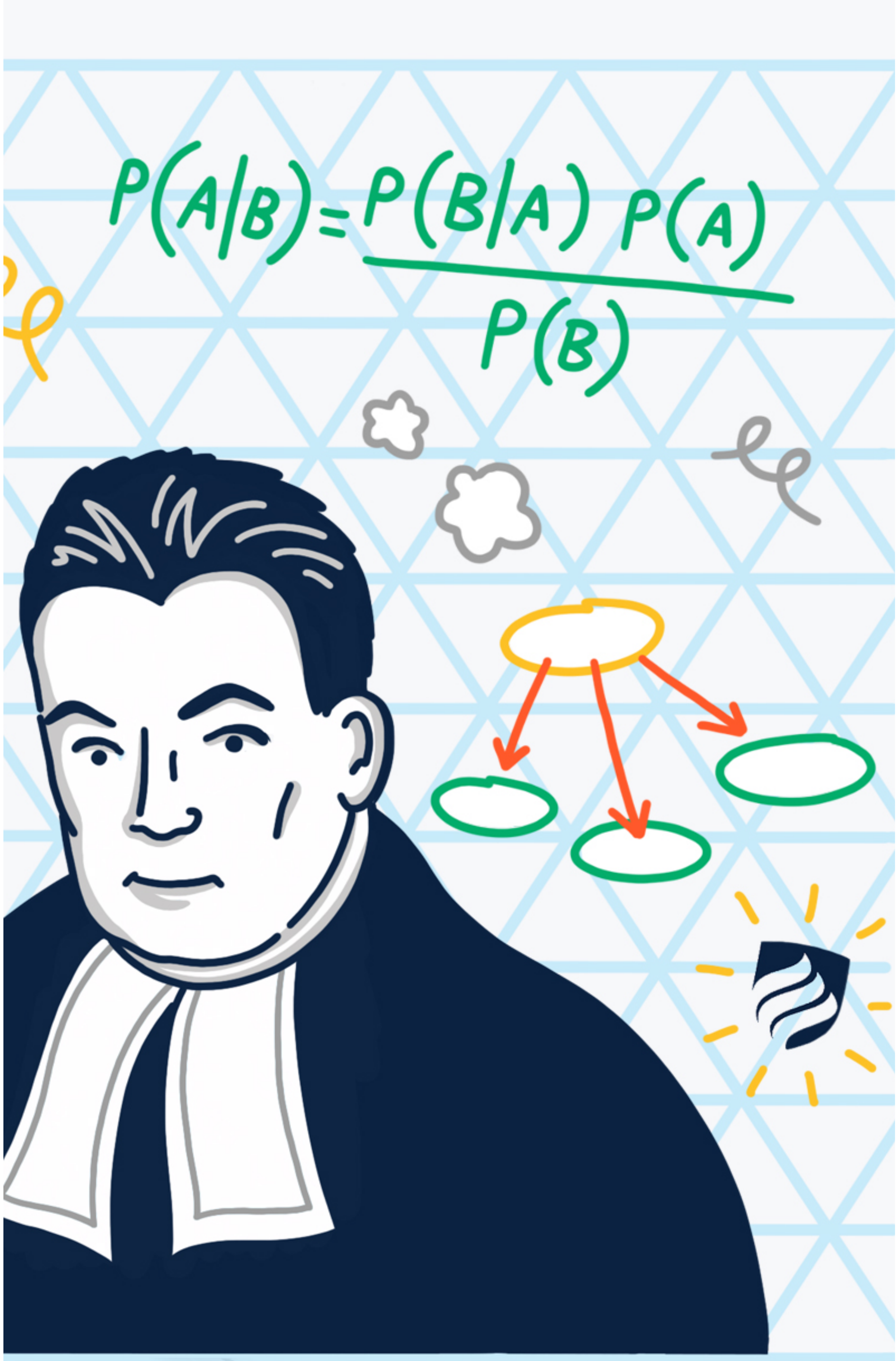
Likelihood: $f(y|p) = (1-p)^y p$

Posterior:

$$\pi(p|\underline{x}) = \pi(p|y) = \frac{f(y|p)\pi(p)}{\underbrace{\boxed{\int_0^1 f(y|p)\pi(p)dp}}_{= c}} = \frac{(1-p)^{y+1}p^2}{c} \Rightarrow p|y \sim Beta(\alpha_1, \beta_1)$$

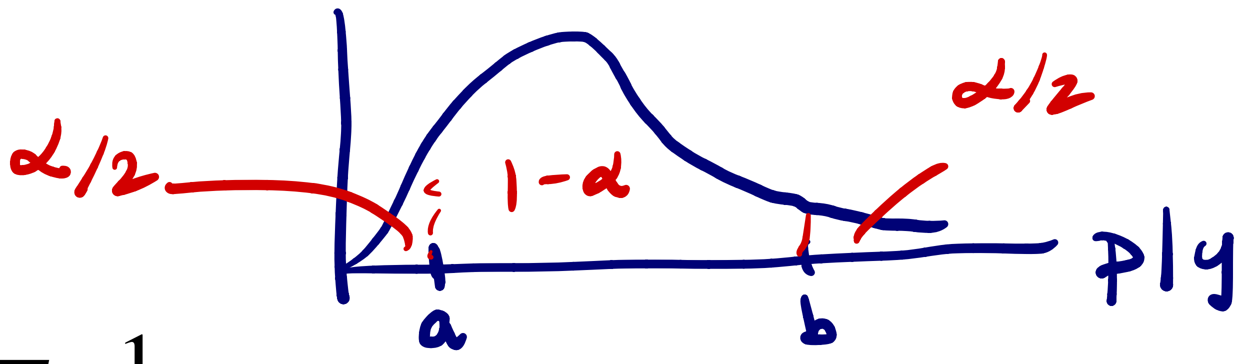
$\alpha_1 = \alpha+1 = 3$

$\beta_1 = y+\beta = y+2$



From: https://bit.ly/3gTjlBg

Example: Suppose researchers would like to know the probability, p, that an individual in a given population has a genetic marker that predisposes them for disease D (this genetic marker is such that an individual has it or doesn't; assume that our test for the marker is completely accurate). Researchers collected data in the following way: they tested people for this genetic marker until they found on person who had it. They stopped data collection at that point. Here's the data: $\mathbf{x} = (0,0,0,0,1)$. Suppose your prior beliefs are best represented by a beta$(2,2)$ distribution. What is the posterior distribution for $p \mid x$?



$95\ \%$ credible interval? Let F^{-1} be the inverse cdf of the posterior distribution for $p \mid x$, i.e., the cdf of beta$(\alpha + 1, y + \beta)$. Then a $95\ \%$ credible interval is given by:

$$\Big(\underbrace{F^{-1}(0.025)}_{a} \, , \, \underbrace{F^{-1}(0.975)}_{b} \Big)$$