Beta Distribution – probability of success, i.e., α

Many times in real life, we come up with scenarios **when we don't know the actual probability but we have "prior knowledge" to guess the probability** (called as prior in Data Science world), beta distribution can be used to **represent all the possible values that probability can take**.

- https://www.kdnuggets.com/2019/09/beta-distribution-what-when-how.html

What is the probability that I get accepted into UCONN CSE Ph.D program? P(Me accepted into Ph.D)

What would be the probability that I get a job at Google after graduation? P(Me accepted to GOOGLE)

What is the probability it will rain tomorrow? P(Rain Tomorrow)

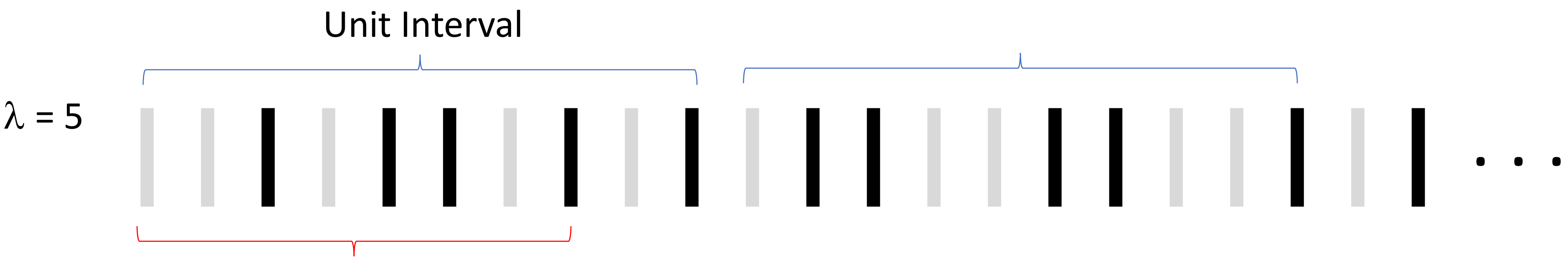
What are the probability of the probability you guessed (i.e., hypothesized)?

Probability: "the extent to which an event is likely to occur" – Oxford Languages

Beta Distribution, given α and β

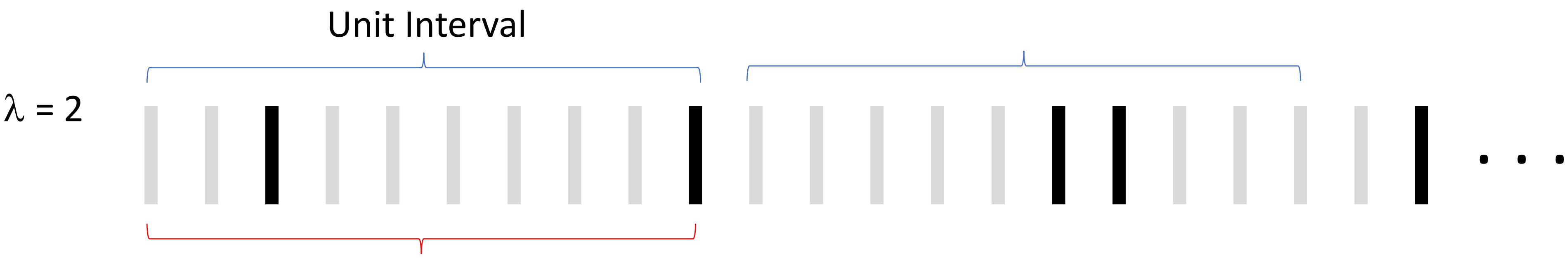
α is to count successes and β to count failures.

Consider a series of events in which hits and misses continue.



k = 1 → After 3 trials.
k = 4 → After 8 trials.

k = 1 → After ? trials.
k = 4 → After ? trials.



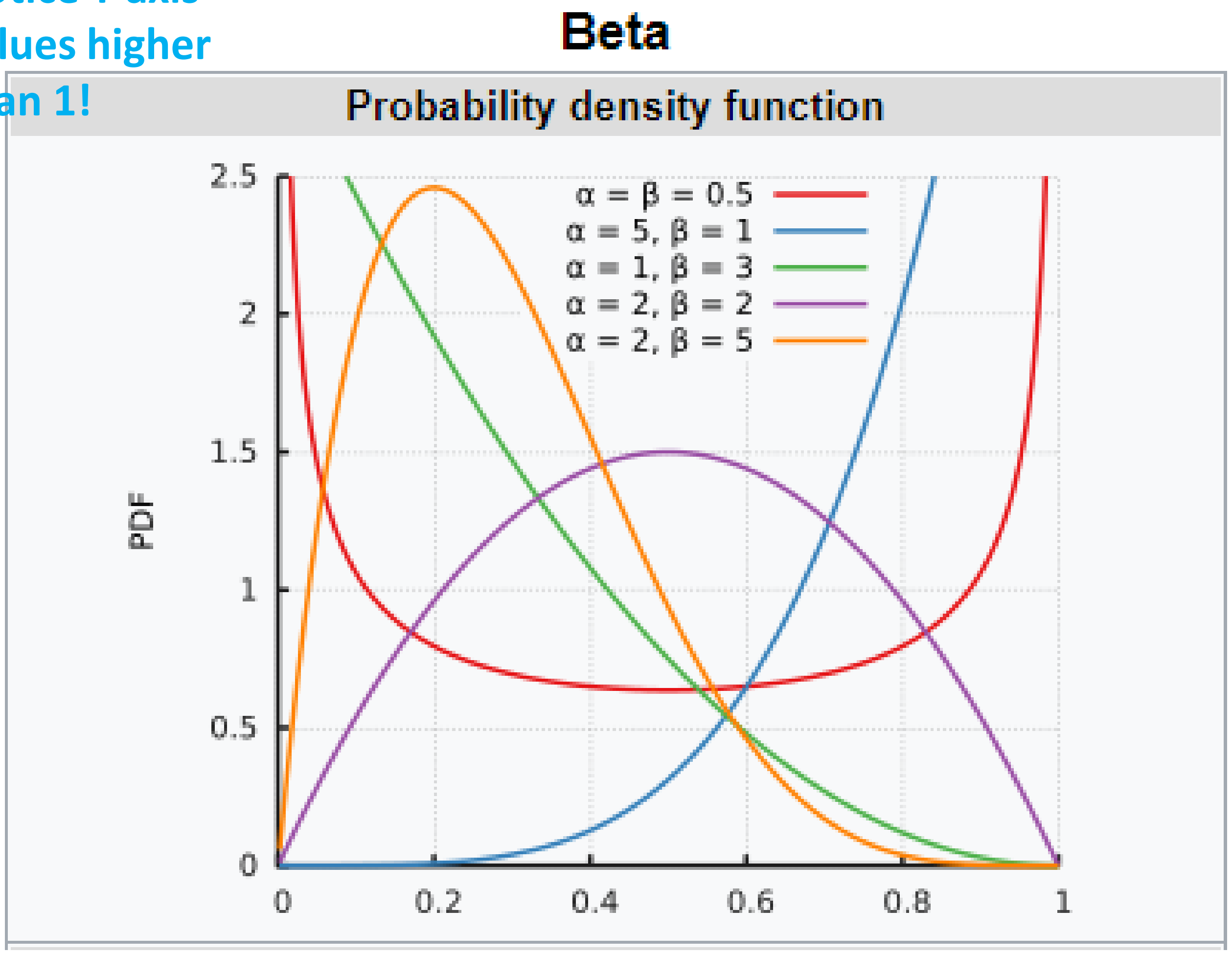
k = 1 → After 3 trials.
k = 2 → After 10 trials.

k = 1 → After ? trials.
k = 2 → After ? trials.

Beta Distribution, given α and β

The beta distribution is a family of **continuous probability distributions defined on the interval [0, 1]** parameterized by two positive **shape parameters, denoted by α and β**, that appear as exponents of the random variable and control the shape of the distribution.

Notice Y axis values higher than 1!



https://en.wikipedia.org/wiki/Beta_distribution

The probability density function (pdf) of the beta distribution, for $0 \le x \le 1$, and shape parameters $\alpha, \beta > 0$,

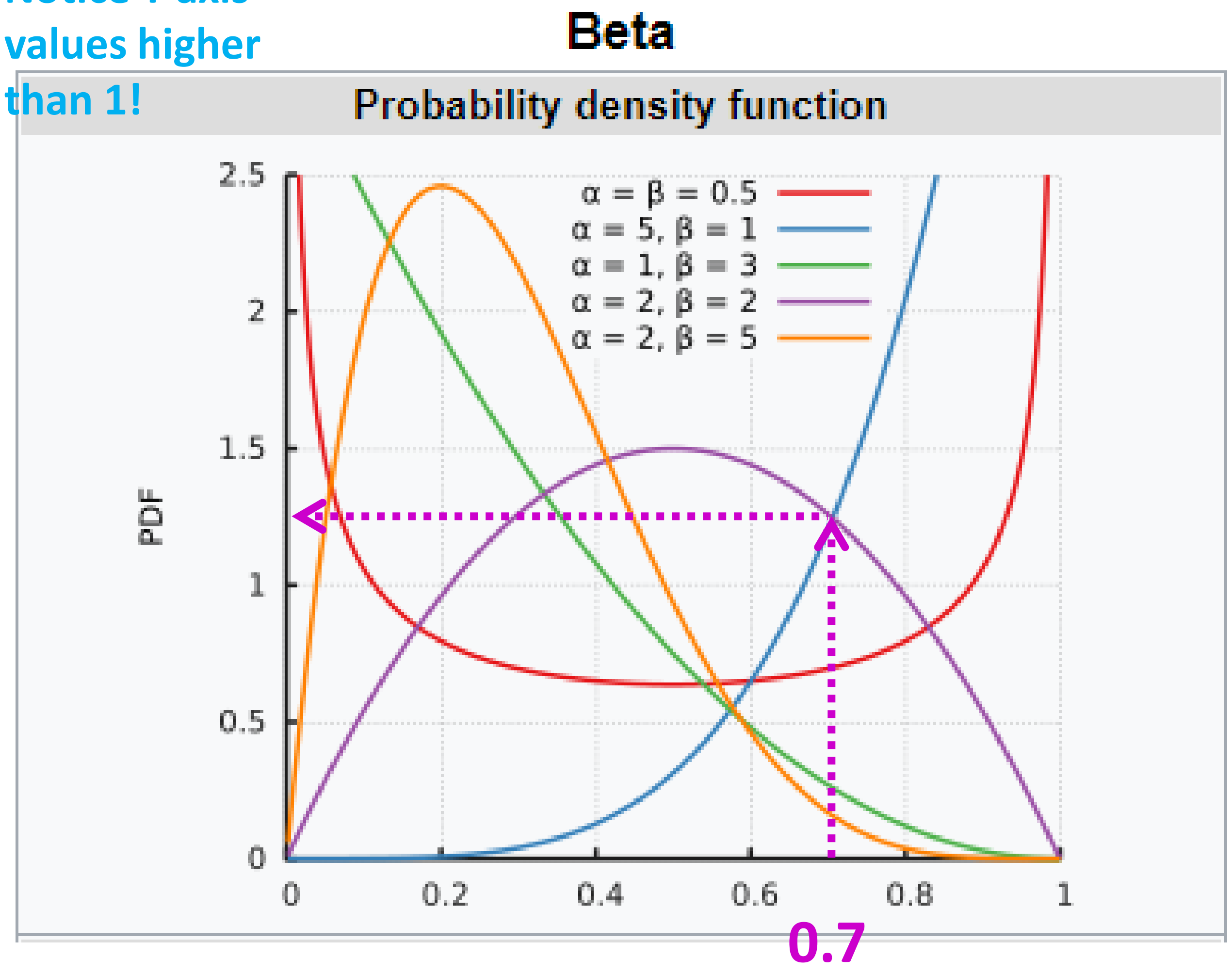
$$\begin{aligned} f(x;\alpha,\beta) &= \mathrm{constant}\cdot x^{\alpha-1}(1-x)^{\beta-1} \\ &= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} \\ &= \underbrace{\boxed{\frac{1}{\mathrm{B}(\alpha,\beta)}}}_{\text{Area}}x^{\alpha-1}(1-x)^{\beta-1} \end{aligned}$$

where Γ(z) is the gamma function. The beta function, B, is a normalization constant to ensure that the total probability is 1. In the above equations x is a realization—an observed value that actually occurred—of a random process X.

Beta Distribution, given α and β

The beta distribution is a family of **continuous probability distributions defined on the interval [0, 1]** parameterized by two positive **shape parameters, denoted by α and β**, that appear as exponents of the random variable and control the shape of the distribution.

Notice Y axis values higher than 1!



$$f(x;\alpha,\beta) = \text{constant} \cdot x^{\alpha-1}(1-x)^{\beta-1}$$

$0.7^{1-1} * (1-0.7)^{1-1}$

$$f(\,0.7;\,\alpha,\,\beta\,)$$

Your guess

Context (player)

Context is defined by # of success/failure.

Four possible values, depending on α, β.

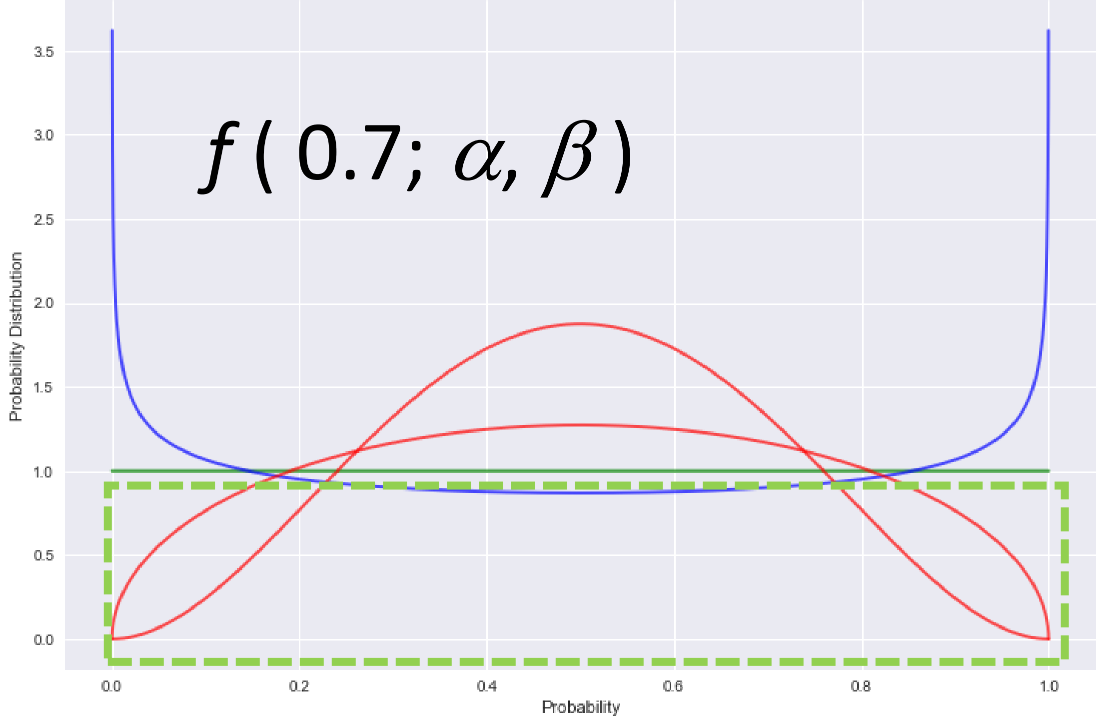
```
ss.beta.pdf(0.7, 1, 1) → 0.9999999999999998
ss.beta.pdf(0.7, 2, 2) → 1.2600000000000005
ss.beta.pdf(0.7, 5, 1) → 0.9999999999999998
ss.beta.pdf(0.7, 0.5, 0.5)→0.6946091180428566
```

Beta Distribution, given α and β

```
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as ss

plt.style.use('seaborn') # pretty  matplotlib plots
plt.rcParams['figure.figsize'] = (12, 8) # what is this for?

def plot_beta(x_range, a, b, mu=0, sigma=1, cdf=False, **kwargs):
    # Plots the f distribution function for a given x if mu and sigma are not
    # provided, standard beta If cdf=True cumulative distribution is plotted.
    # Passes any keyword arguments to matplotlib plot

    x = x_range
    if cdf:
        y = ss.beta.cdf(x, a, b, mu, sigma)
    else:
        y = ss.beta.pdf(x, a, b, mu, sigma)
    plt.plot(x, y, **kwargs)

x = np.linspace(0, 1, 5000)
plot_beta(x, -0.0+1, -0.0+1, 0, 1, color='green', lw=2, ls='-', alpha=0.7) #0/0
plot_beta(x, -0.2+1, -0.2+1, 0, 1, color='blue', lw=2, ls='-', alpha=0.7) #0.2/0.2
plot_beta(x, 0.5+1, 0.5+1, 0, 1, color='red', lw=2, ls='-', alpha=0.7) #0.5/0.5
plot_beta(x, 2+1, 2+1, 0, 1, color='red', lw=2, ls='-', alpha=0.7) #2/2
plt.ylabel('Probability Distribution')
plt.xlabel ('Probability')
plt.show()
```

ss.beta.pdf(0.7, 1, 1) → 0.9999999999999998

All cases, mean = 0.5 and symmetric.



When $\alpha = \beta = 1$, it becomes uniform, i.e., 1 regardless of x.

$f(0.7; 1, 1) = 1.$

Beta Distribution, given α and β

Beta distribution is widely used to **model the prior beliefs** or probability distribution in real world applications.

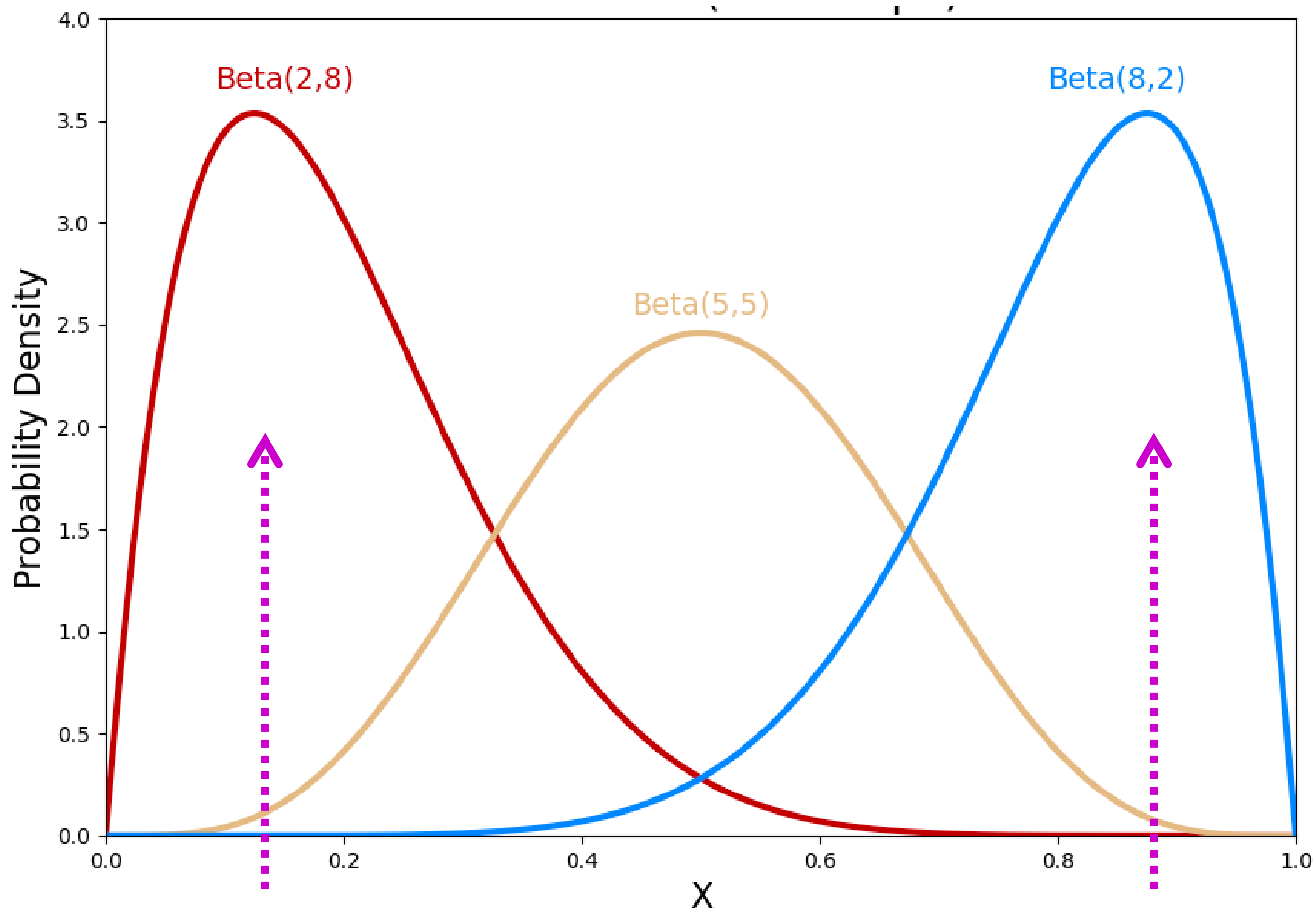


Fig 1. Change in Prior Distributions (Beta Distribution)

https://vitalflux.com/beta-distribution-explained-with-python-examples/

$$f(x;\alpha,\beta) = \text{constant}\cdot x^{\alpha-1}(1-x)^{\beta-1}$$
$$= \frac{1}{\mathrm{B}(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

$$\mathrm{E}[X] = \frac{\alpha}{\alpha+\beta}$$

$$\mathrm{var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Make $\alpha > \beta$ and see how the shape changes.
Make $\alpha < \beta$ and see how the shape changes.

If α increases, will the peak be shifted to right or left? → polling

Beta Distribution, given α and β

Beta distribution is widely used to **model the prior beliefs** or probability distribution in real world applications.

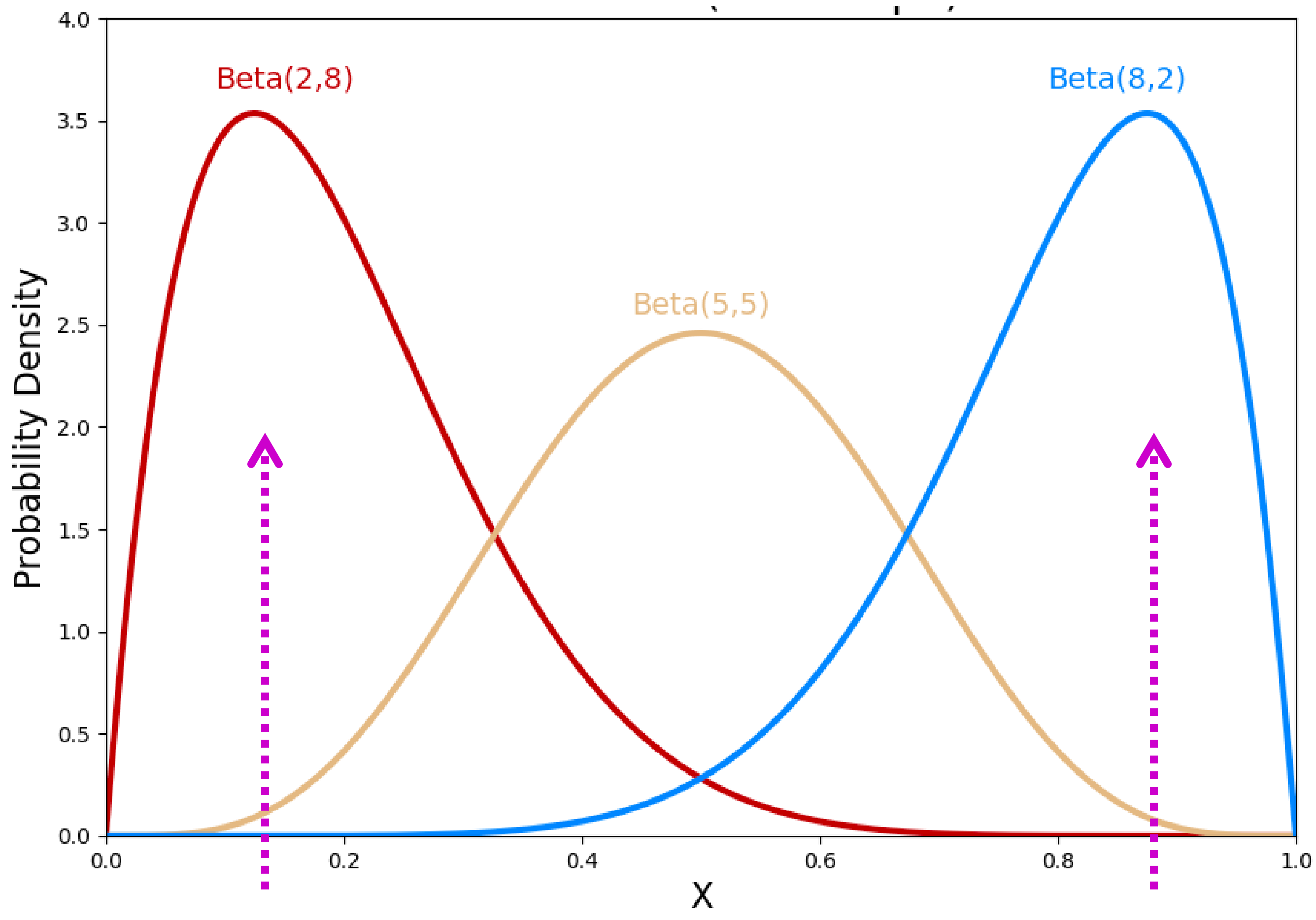


Fig 1. Change in Prior Distributions (Beta Distribution)

$$f_X(x;q,r) = \begin{cases} \dfrac{1}{B(q,r)} \dfrac{(x-a)^{q-1}(b-x)^{r-1}}{(b-a)^{q+r-1}}, & \text{for } a \le x \le b \\ =0, & \text{otherwise} \end{cases}$$

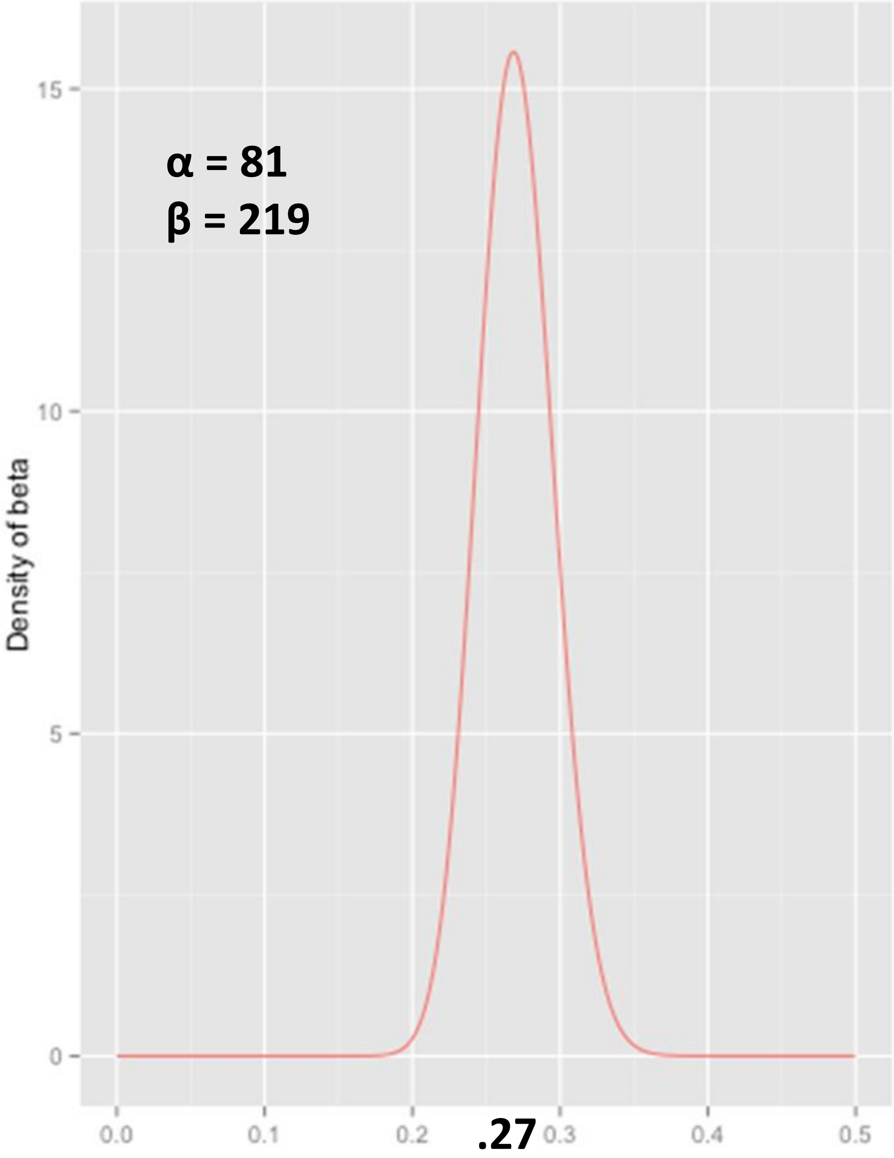
Even if the baseball player got strikeout in first couple of matches, one still may chose to believe based on his prior belief (prior distribution) that he would end up achieving his batting average.

However, as the match proceed, the prior may change based on his performance in subsequent matches. This would mean altering the parameters value of α and β.

https://vitalflux.com/beta-distribution-explained-with-python-examples/

Understanding the beta distribution (using baseball statistics)

The idea of a probability distribution of probabilities means **generating all the possible values of a probability since you do not know what that correct probability is** (e.g., batting average).



Batting averages

Batting averages— The number of times a player gets a base hit divided by the number of times he goes up at bat (so it's just a percentage between 0 and 1). .266 → an average batting average; .300 → an excellent one.

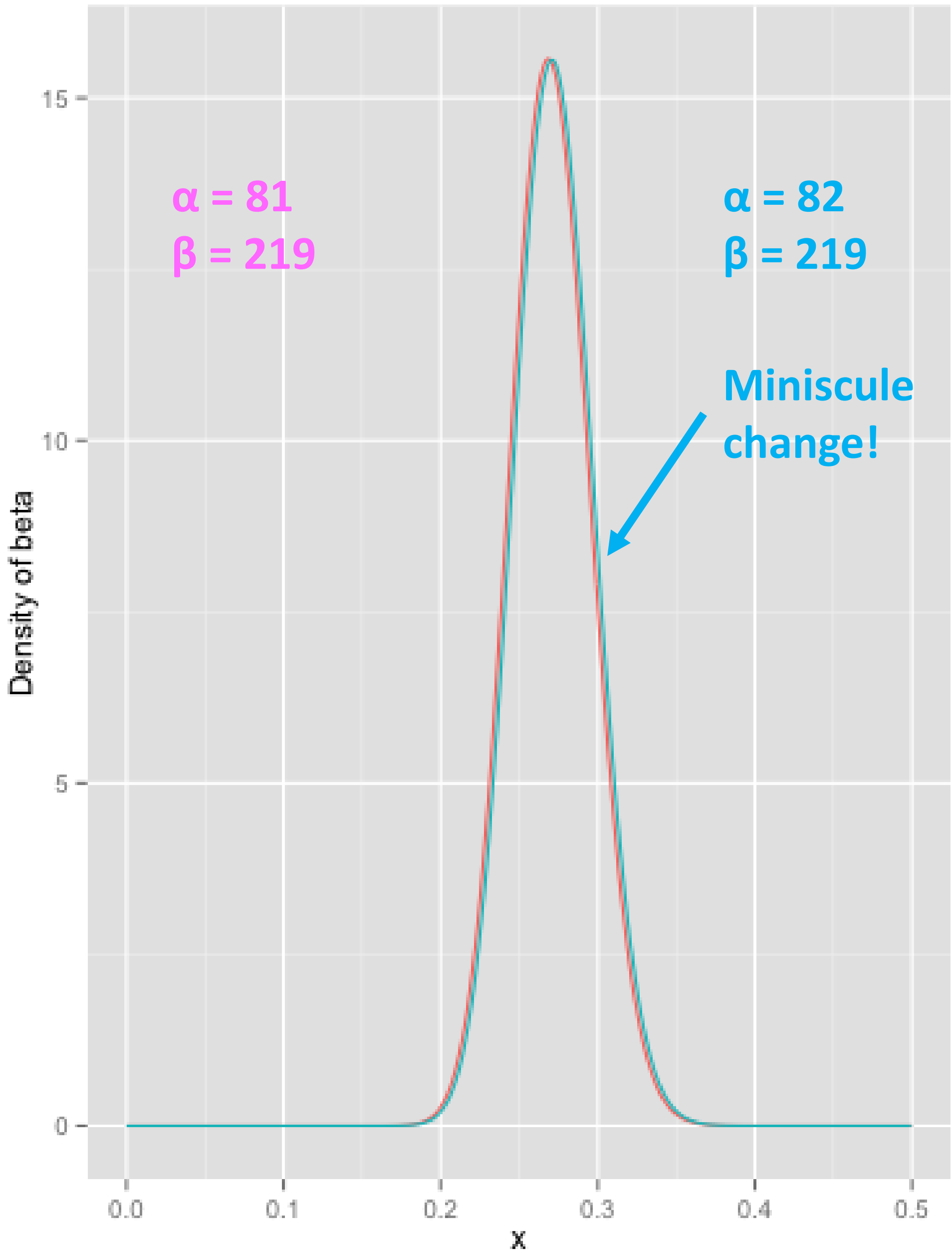
We expect that the player's season-long batting average will be most likely around .27, but that it could reasonably range from **.21 to .35.** This can be represented with a beta distribution with parameters α=81 and β=219:

$$\frac{1}{\mathrm{B}(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

- The mean is $\frac{\alpha}{\alpha+\beta} = \frac{81}{81+219} = .270$
- As you can see in the plot, this distribution lies almost entirely within $(.2, .35)$ - the reasonable range for a batting average.

http://varianceexplained.org/statistics/beta_distribution_and_baseball/

Understanding the beta distribution (using baseball statistics)



Joe's last season batting average was .27 with 81 hits and 219 misses.

mu = (α / α + β) = 81/ (81 +219) = 0.27

Joe gets a hit on the first day at his first bat.

His record for the season is now "1 hit; 1 at bat."

What would be the batting average at the end of the season?

→ Should it be 1.0?

Update what you already know!

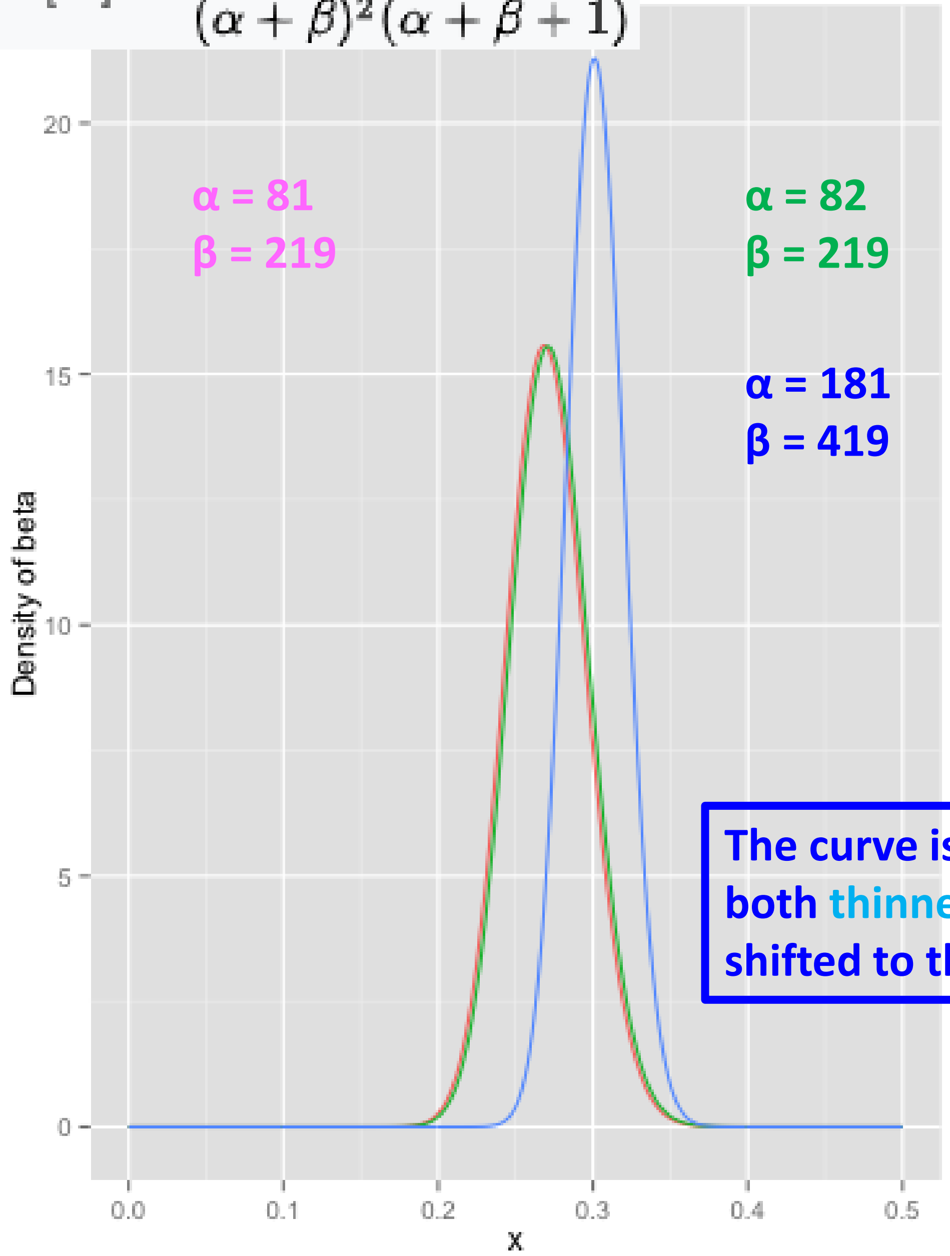
The new beta distribution will be:

Beta(α_0 + hits, β_0 + misses)

Beta(81,219) → Beta(81+1,219)

Understanding the beta distribution (using baseball statistics)

$$\mathrm{var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$



The curve is now both thinner and shifted to the right.

Joe's last season batting average was .27 with 81 hits and 219 misses.

mu = (α / α + β) = 81/ (81 +219) = 0.270

Joe gets a hit on the first day at his first bat.

mu_1 = (α / α + β) = 82/ (82 +219) = 0.272

Halfway season, at 300 bats, his record is 100 hits and 200 misses.

Beta(81,219) → Beta(81+100, 219+200)

mu_{300} = (α / α + β) = 181 / (181 + 419) = 0.301

Update what you already know!

mu_{300*} = (α / α + β) = 100 / (100 + 200) = 0.333

Should Joe's seasonal batting average be 0.333?