

Probability and Babies

This demo used to be live. We now know that the delivery happened on Jan 23rd. Lets go back in time to Jan 1st and see what the probability looked like at that point.

What is the probability that Laura gives birth today (given that she hasn't given birth up until today)?

Today's Date	1/Jan/2021
--------------	------------

Due Date	18/Jan/2021
----------	-------------

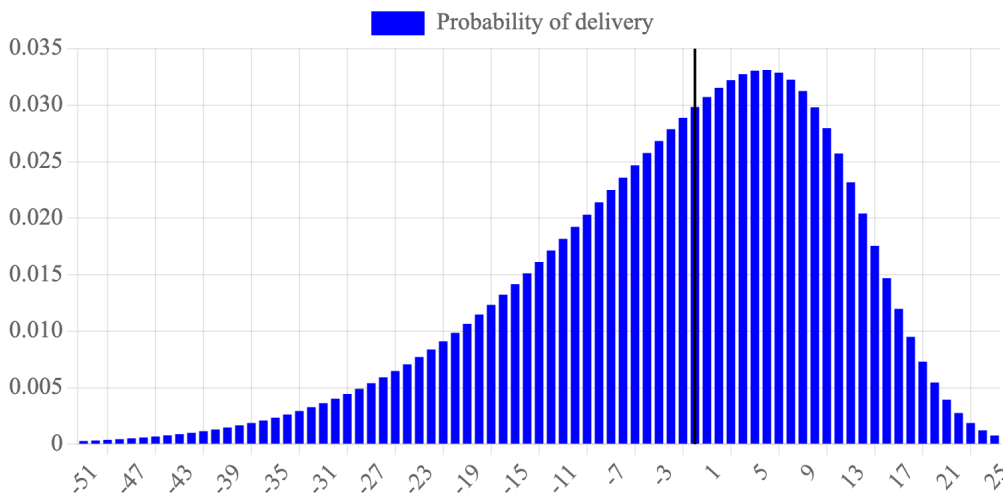
Probability of delivery today: **0.014**

Probability of delivery in next 7 days: **0.144**

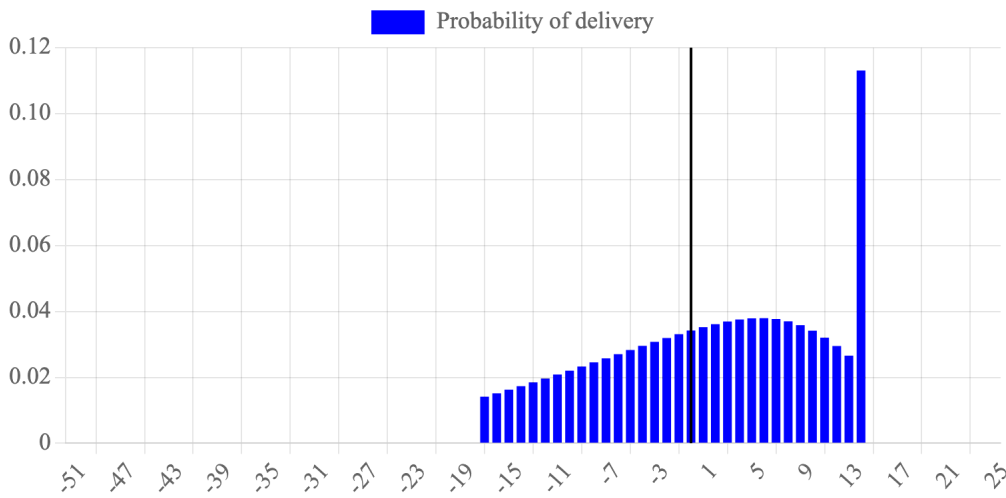
Current days past due date: **-17** days

Unconditioned probability mass before today: **0.128**

How likely is delivery, in humans, relative to the due date? There have been millions of births which gives us a relatively good picture [1]. The length of human pregnancy varies by quite a lot! Have you heard that it is 9 months? That is a rough, point estimate. The mean duration of pregnancy is 278.6 days, and pregnancy length has a standard deviation (SD) of 12.5 days. This distribution is not normal, but roughly matches a "skewed normal". This is a general probability mass function for the first pregnancy collected from hundreds of thousands of women (this PMF is very similar across demographics, but changes based on whether the woman has given birth before):



Of course, we have more information. Specifically, we know that Laura **hasn't** given birth up until today (we will update this example when that changes). We also know that babies which are over 14 days late are "[induced](#)" on day 14. How likely is delivery given that we haven't delivered up until today? Note that the y-axis is scaled differently:



Let's approach this problem formally using inference. First we introduce a random variable D to represent the days after delivery that the baby is born. Note that D can be negative if the baby is due before the due date. We can use Inference to update our belief in D given our observation that we have not delivered yet:

$$P(D = i | \text{No Baby Yet}) = \frac{P(\text{No Baby Yet} | D = i) P(D = i)}{P(\text{No Baby Yet})}$$

$P(\text{No Baby Yet} | D = i)$ is always either 1 or 0. Note that conditioning on $D = i$ means we are being told the actual date of delivery. If the delivery hasn't happened yet (eg today is before i) then the probability of No Baby Yet is 1. If the delivery has already happened (eg today is after i) then the probability of No Baby Yet is 0.

$P(D = i)$ is our prior belief (the probability of delivery dates based off historical data). $P(\text{No Baby Yet})$ is the normalization constant. Instead of calculating it explicitly, we can compute the numerators for each value of i . We can then normalize the distribution (compute the sum of the numerators, and divide every probability by this sum) in order to implicitly compute it. An equivalent (but more compute heavy) solution would be to expand $P(\text{No Baby Yet})$ using the law of total probability:

$$P(\text{No Baby Yet}) = \sum_i P(\text{No Baby Yet} | D = i) P(D = i)$$

How do we deal with the fact that babies are induced after $D = 14$? Well we can adjust our prior so that all of the probability for days 14 and on is shifted to day 14. This is equivalent to the following calculation:

$$P(D = 14) = \sum_{i \geq 14} P(D = i)$$

```

def update_belief_baby(prior, today = -19):
    # pr_D[i] is P(D = i | No Baby Yet).
    pr_D = {}
    min_i = -50
    max_i = 14
    for i in range(min_i, max_i + 1):
        # P(NoBaby | D = i)
        likelihood = 0 if i < today else 1
        pr_D[i] = likelihood * prior[i]
    # implicitly computes the LOTP
    normalize(pr_D)
    return pr_D

def normalize(unnormalized_pmf):
    total_sum = sum(unnormalized_pmf.values())
    normalized = {}
    for key, value in unnormalized_pmf.items():
        normalized[key] = value / total_sum
    return normalized

```

Extension Problem

Chris had two other good friends who had babies with the exact same due date (Really! This actually happened). What is the probability that all three babies are delivered on the exact same day?

Probability of three couples on the same day: **0.002**

How did we get that number? Let p_i be the probability that one baby is delivered on day i -- this number can be read off the probability mass function. Let D_i be the event that all three babies are delivered on day i . Note that the event D_i is [mutually exclusive](#) with the event that all three babies are born on another day (So for example, D_1 is mutually exclusive with D_2, D_3 etc). Let $N = 3$ be the event that all babies are born on the same day:

$$\begin{aligned}
 P(N = 3) &= \sum_i P(D_i) && \text{Since days are mutually exclusive} \\
 &= \sum_i p_i^3 && \text{Since the three couples are independent}
 \end{aligned}$$

[1] [Predicting delivery date by ultrasound and last menstrual period in early gestation](#)

Acknowledgements: This problem was first posed to me by Chris Gregg.