Inferring age distribution of blood cells

Aodong Li
Computer Science Department
New York University
New York, NY 10003
al5350@nyu.edu

Abstract

HbA1c is a measurement of the glycation percentage of hemoglobin – a specific kind of protein in blood cells. Glycation of proteins results from the chronic attachment of glucose in the bloodtream. It turns out that the measurement is highly correlated to the glucose concentration over the preceding 8–12 weeks and to the blood cell age distribution. The relationship between HbA1c and the average glucose concentration has been rigorously investigated by Nathan et al. [7]. However, much detailed connection among HbA1c, varying glucose level, and the blood cell age distribution has not been touched yet. In this paper, we sought to define the mathematical relationship among these three factors and investigate the assumptions under which the inference of the blood cell age distribution is feasible. In particular, we define their connections as a convolution operation. Simulation results show that the proposed assumptions are consistent if the cell age distribution is uniform. Other theoretical aspects involving how to distinguish the uniform distribution and the finite exponential distribution, how to specify a linear programming problem for inferring a generic distribution, and the possibility of applying the pointwise deconvolution are also investigated.

1 Introduction

Blood cells in the bloodstream are exposed to varying levels of sugar (glucose, $G(t)$). For healthy people, fingerstick capillary glucose concentration has a mean of about 100 mg/dL with a standard deviation of about 10 mg/dL. For people with type I diabetes, mean fingerstick glucose can be as high as 250 mg/dL, and the standard deviation can be as high as 60 mg/dL.

Due to the affinity of the molecular structures, glucose is automatically attached to some particular part of the proteins in blood cells. These proteins become glycated irreversibly at a rate that depends on $G(t)$. The glycation process is slow, and even the oldest cells rarely reach a 20% level of glycation. The glycation process is never saturated.

We can take a blood sample and measure the overall glycated fraction of proteins in all the cells. Blood cells contain various proteins. If the target protein is hemoglobin, its glycated form is called hemoglobin A1c, and the measurement is referred to as HbA1c.

Because the blood cells are in circulation for approximately 120 days and the glycation is accumulated in the cells over this period, HbA1c has been standardized to monitor the long-term glycemic control according to Berg and Sacks [1]. Moreover, this accumulation process makes the concentration of HbA1c depend on both the glucose concentration $G(t)$ and the lifespan of blood cells.

Nathan et al. [7][8][6] have studied rigorously the relationship between HbA1c and the average glucose concentration over the preceding 8–12 weeks, and confirmed the striking linear connection. To our knowledge, however, the precise interplay between HbA1c, $G(t)$, and the cell lifespan (or the age distribution $f_{\mathrm{RBC_{age}}}(t)$) has not been investigated in a much detailed manner.

In this paper, we sought to explore the precise mathematical relationship among HbA1c, $G(t)$, and $f_{\text{RBC}_{\text{age}}}(t)$ whilst establishing the necessary assumptions to make the inference of $f_{\text{RBC}_{\text{age}}}(t)$ possible. In summary, we make the following contributions:

1. We define the mathematical relationship among these factors as a convolution operation. In particular, HbA1c is a convolution of $G(t)$ and $f_{\text{RBC}_{\text{age}}}(t)$.

2. When $f_{\text{RBC}_{\text{age}}}$ is uniform, we show that it can be inferred easily with one measurement HbA1c.

3. Based on the previous results of Nathan et al. [7], Kahn et al. [5], the convolution computation is refined such that the results are more consistent at least if $f_{\text{RBC}_{\text{age}}}(t)$ assumes a uniform distribution.

4. We specify an easily achieved sequence $G(t)$ in clinical environment to distinguish $f_{\text{RBC}_{\text{age}}}(t)$ as a uniform distribution from a finite exponential distribution.

5. Without assumptions on the shape of $f_{\text{RBC}_{\text{age}}}(t)$, we sought to specify a linear programming optimization problem to infer $f_{\text{RBC}_{\text{age}}}(t)$ and explore the possibility of applying the pointwise deconvolution to perform the inferrence.

2 Related work

Previous work by Nathan et al. [7] on figuring out the relation between average glucose (AG) concentration and HbA1c adopted the linear regression analysis. And they provided the tightest correlation:

$$\text{HbA1c}_{\%} = \frac{\text{AG}_{\text{mg/dl}} + 46.7}{28.7}, R^2 = 0.84, P < 0.0001 \tag{1}$$

where the subscript indicates the unit. Besides, this work provides important hints on refining our assumptions and computations in that *the occurrence of the bias term indicates HbA1c is not zero even if glucose concentration is zero*.

Furthermore, clinical information from Nathan et al. [6] and Goldstein [3] show that HbA1c is also correlated with the blood cell age distribution. Although the HbA1c reflects mean blood glucose over the entire 120-day lifespan of the red blood cell, it correlates best with blood glucose over the previous 8 to 12 weeks.

Hence incorporating the role of blood cell age distribution is necessary to provide a more precise description of HbA1c and we show that under certain assumptions, inferring the age distribution is feasible.

3 Assumptions

In this section, we give the necessary assumptions and define the mathematical relationship between HbA1c, glucose concentration $G(t)$ and the cell age distribution $f_{\text{RBC}_{\text{age}}}(t)$.

We denote t_0 as the current time and $t < t_0$ as preceding time. HbA1c(t_0) and $G(t_0)$ are the current measurements and HbA1c$(t), G(t)$ for $t < t_0$ are the previous measurements. Here comes the first assumption.

- Assumption 1: $f_{\text{RBC}_{\text{age}}}$ remains constant over the time that measurements take place.

In other words, we suppose the varying glucose concentration does not affect the age of the cells as well as the birth and the death of the cells, so that we can measure the age distribution $f_{\text{RBC}_{\text{age}}}$ accurately.

The accumulation of glucose in the cells happens as follows. $G(t_0)$ affects the whole cells in the blood for a very short period of time Δt. However, preceding glucose concentration $G(t - \Delta t)$ cannot affect the current whole cells because cells are died and born continuously, and the new cells born from $t - \Delta t$ to t are not influenced by glucose concentration $G(t - \Delta t)$. In fact, $G(t - \Delta t)$ affects the survived cells with age at least Δt, and $G(t - 2\Delta t)$ affects the survived cells with age at

least $2\Delta t$, etc. This accumulation process results in

$$\text{HbA1c}(t) \propto \sum_{n=0}^{\infty} G(t - n\Delta t)F(\geq n\Delta t)\Delta t$$

where $F(\geq n\Delta t)$ is the fraction of cells whose age is at least $n\Delta t$. Note that the right-hand side is bounded since the cell cannot survive forever. In other words, $\exists N > 0, \forall n \geq N, F(\geq n\Delta t) = 0$. If we denote $f_{\text{RBC}_{\text{age}}}$ as the probability density function for the distribution, then

$$F(\geq n\Delta t) = 1 - \int_0^{n\Delta t} f_{\text{RBC}_{\text{age}}}(t)\mathrm{d}t.$$

Now we take the limiting case as $\Delta t \to 0$, the summation becomes integral and the relationship becomes

$$\text{HbA1c}(t) \propto \int_0^{\infty} G(t - a)\left(1 - \int_0^a f_{\text{RBC}_{\text{age}}}(\tau)\mathrm{d}\tau\right)\mathrm{d}a. \tag{2}$$

In order to make the above relation an equation, we need other assumptions to specify the glycation rate with respect to $G(t)$.

Define the glycation rate $v(t) = f(G(t))$ – the added percentage of glycation per day. Specifically, one of its valid units can be percent/day, which corresponds to the percentage unit of HbA1c.

From the background information in section 1, we further consider the following assumptions:

- Assumption 2: suppose the glucose concentration G contributes linearly to the glycation rate v.
- Assumption 3: suppose the mean lifespan of the cells is 120 days.
- Assumption 4: suppose the oldest cells reach a 20% level of glycation in 250 mg/dL glucose concentration, which is the inner environment of a person with type I diabetes.

Based on the above assumptions, we can readily write the glycation rate $v(t)$ as a function of the glucose concentration $G(t)$:

$$v(t) = \frac{G(t)20\%}{250\text{mg/dL} \times 120\text{d}} = c \cdot G(t).$$

We call c the glycation constant.

Once we specify the glycation rate $v(t)$, we can formulate the equation to compute HbA1c

$$\text{HbA1c}(t) = \int_0^{\infty} \frac{G(t - a)20\%}{250\text{mg/dL} \times 120\text{d}}\left(1 - \int_0^a f_{\text{RBC}_{\text{age}}}(\tau)\mathrm{d}\tau\right)\mathrm{d}a, \tag{3}$$

which coincides in the form of convolution operation.

In addition, one more thing needs to be specified. Notice that from equation 1, the occurrence of the bias term 1.627 indicates that even without glucose concentration, HbA1c is not assumed to be zero. From this knowledge, equation 3 should be refined by adding an extra bias,

$$\text{HbA1c}(t) = \int_0^{\infty} \frac{G(t - a)20\%}{250\text{mg/dL} \times 120\text{d}}\left(1 - \int_0^a f_{\text{RBC}_{\text{age}}}(\tau)\mathrm{d}\tau\right)\mathrm{d}a + 1.627. \tag{4}$$

The bias term turns to be important as we will see in the experiment section that this bias renders the inference results much more consistent because we use equation 1 to compute HbA1c and this gives the tightest approximation to the true measurement.

4 Methods

In this section, we state some techniques to infer $f_{\text{RBC}_{\text{age}}}$. It turns out if we assume the shape of the distribution, i.e., the parametric family of the distribution by finite parameters, we can solve the parameters quickly. Otherwise, we have to specify the support size to infer the shape of the density function by linear programming.

In practice, we have the access to the discrete sampling sequence of glucose concentration $G(n)$ with some fixed interval, and in the absence of HbA1cmeasurement we can use equation 1 to compute it. The methods are all in time domain and we approximate the integral in equation 4 by Riemann sum.

4.1 Uniform distribution

Suppose $f_{\text{RBC}_{\text{age}}}$ conforms to some uniform distribution with mean M_{RBC},

$$f_{\text{RBC}_{\text{age}}}(t) = \frac{1}{2M_{\text{RBC}}}, \qquad 0 \le t \le 2M_{\text{RBC}},$$

and the equation 4 can be re-written as

$$\text{HbA1c}(t) = \int_0^{2M_{\text{RBC}}} v(t-a) \left(\frac{2M_{\text{RBC}} - a}{2M_{\text{RBC}}} \right) \mathrm{d}a + 1.627.$$

For this convolution, there is only one parameter that needs to specify, thus only one measurement of HbA1c is required. However, it is hard to write out the formula of M_{RBC} directly, so we will use the numerical way to approximate M_{RBC}.

Here we utilize the Riemann sum to approximate the integral,

$$\text{HbA1c} - 1.627 = \sum_{n=0}^{(2M_{RBC}-1)/\Delta t} v(-n) \times \Delta t \times \frac{2M_{RBC} - n \cdot \Delta t}{2M_{RBC}} \tag{5}$$

where $v(-n) = c \cdot G(-n)$ and we denote the negative sequence $G(-n)$ as the reverse order of $G(n)$ where $G(0)$ is the last entry and $G(-1)$ is the second last entry. The same semantics apply to the glycation rate $v(n)$, and also to the following sections.

Equation 5 as well as equation 1 render us the ability to compute M_{RBC} in practice.

4.2 Finite exponential distribution

Instead of uniform distribution, suppose $f_{\text{RBC}_{\text{age}}}$ conforms to finite exponential distribution, with parameter M_{RBC} and $k_r \sim \frac{1}{200 M_{\text{RBC}}}$,

$$f_{\text{RBC}_{\text{age}}}(t) = \frac{\exp(-k_r \cdot t)}{\frac{1}{k_r}(1 - \exp(-k_r \cdot 2 \cdot M_{\text{RBC}}))}, \qquad 0 \le t \le 2M_{\text{RBC}},$$

and in this case the equation 4 can be re-written as

$$\text{HbA1c}(t) = \int_0^{2M_{\text{RBC}}} v(t-a) \left(\frac{\exp(-k_r \cdot a) - \exp(-k_r \cdot 2 \cdot M_{\text{RBC}})}{1 - \exp(-k_r \cdot 2 \cdot M_{\text{RBC}})} \right) \mathrm{d}a + 1.627.$$

Also the Riemann sum approximation is

$$\text{HbA1c} - 1.627 = \sum_{n=0}^{(2M_{\text{RBC}}-1)/\Delta t} v(-n) \times \Delta t \times \left(\frac{\exp(-k_r \cdot n \cdot \Delta t) - \exp(-k_r \cdot 2 \cdot M_{\text{RBC}})}{1 - \exp(-k_r \cdot 2 \cdot M_{\text{RBC}})} \right). \tag{6}$$

If we can make at least two different measurements of HbA1c, then in theory we can solve this problem because there are two variables and two constraints. Experiments for this section shall be taken in the future.

4.3 Distinguishing uniform distribution from finite exponential distribution

If we can measure multiple HbA1c and glucose concentration, we might be able to discern the difference between a uniform distribution and a finite exponential distribution, due to their very different density shapes. The outputs, HbA1c will finally reflect the density function difference if there are perturbations in the inputs G.

Many sequences of G and HbA1c can distinguish the distributions, such as linearly increasing G, step-like G, pulse G, etc. Here we only consider step-like G and pulse-like G for practical simplicity and realization. (We can find such sequences easily in clinical environment.) Their plots are shown in figure 1. Now denote the HbA1c as $H(0)$ before G increases and denote the HbA1c after G increases as $H(n)$, $n > 0$. The index corresponds to $G(n)$. For the step-like sequence G, the output HbA1c
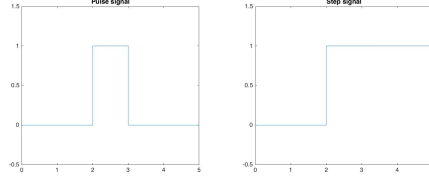
Figure 1: Pulse-like and step-like signal.

should increase with a slower and slower rate as time goes, since $(H(n) - H(n-1)) \propto F(\geq n\Delta t)$, which means that the increment difference between neighboring HbA1c measurements should be proportional to the fraction of the cells that do not have the same glycation rate in their whole life, i.e., those cells whose age is at least $n\Delta t$. Thus the increment difference should be linear for uniform distribution and exponential for finite exponential distribution.

If we use pulse-like sequence G, the output HbA1c should be expected to increase at the pulse and then decrease afterwards due to resetting the glucose concentration. Specifically, $H(n) \propto F(\geq n\Delta t)$, which means HbA1c is proportional to the fraction of cells that have been influenced by the impulse. The fraction is decreasing because cells are subject to death constantly. Thus $H(n)$ should decrease linearly for uniform distribution and exponentially for finite exponential distribution.

4.4 Linear programming for generic distribution

As soon as we represent the integral as Riemann sum, we can denote the problem by linear programming. To simplify the notation, as in the above section, let $H(t)$ be HbA1c(t) and $F(t)$ be $1 - \tilde{F}_{\text{RBC}_{\text{age}}}(t; \theta)$ where $\tilde{F}_{\text{RBC}_{\text{age}}}(t; \theta) = \int_0^t f_{\text{RBC}_{\text{age}}}(\tau)\mathrm{d}\tau$ is the cumulative distribution function of the blood cell age distribution, which is controlled by the parameter θ. At this moment, conforming to the previous assumptions, we assume that θ does not change over time and is an unknown constant.

Now suppose we can measure multiple Hs and Gs. The best case would be one measurement of G and there is one corresponding measurement of H at the same time.

The following equations is an approximation of equation 4.

$$\tilde{H}(0) = \frac{H(0) - 1.627}{c \cdot \Delta t} = G(0)F(0) + G(-1)F(1) + ... + G(-n)F(n)$$
$$\tilde{H}(1) = \frac{H(1) - 1.627}{c \cdot \Delta t} = G(1)F(0) + G(0)F(1) + ... + G(-(n-1))F(n)$$
$$...$$
$$\tilde{H}(n) = \frac{H(n) - 1.627}{c \cdot \Delta t} = G(n)F(0) + G(n-1)F(1) + ... + G(0)F(n)$$

which is essentially a linear system and is equivalent to the matrix form,

$$\tilde{G} \cdot F = \tilde{H}$$

where $1 = F(0) \geq F(1) \geq ... \geq F(n) \geq 0$ performs as the constraint in order for F to be a valid cumulative distribution function.

The only problem left is to specify the support size n of the age distribution. So we can construct a linear programming (LP) with the objective function to minimize the distance between $F(n)$ and 0. But before reaching the final optimization problem, the constraints need to be represented in matrix form.

The constraint $1 = F(0) \geq F(1) \geq ... \geq F(n) \geq 0$ can be rewritten as

$$F(0) = 1$$
$$F(0) - F(1) \geq 0$$
$$F(1) - F(2) \geq 0$$
$$...$$
$$F(n-1) - F(n) \geq 0$$
$$F(n) \geq 0$$

which is equivalent to the form $AF \geq 0$ and $F(0) = 1$ where A is a band matrix

$$A = \begin{bmatrix} 1 & -1 & 0 & ... & 0 \\ 0 & 1 & -1 & 0 & ... \\ ... & & & & \\ 0 & 0 & ... & 0 & 1 \end{bmatrix}.$$

The final constrained problem is

$$\min_F \left(c^\top F \right)$$
$$\text{subject to } \tilde{G}F = \tilde{H}$$
$$AF \succeq 0$$
$$F(0) = 1$$

where $c = (0, ..., 0, 1)^\top \in \mathbf{R}^n$, $A \in \mathbf{R}^{n \times n}$, $\tilde{G} \in \mathbf{R}^{n \times n}$, $F \in \mathbf{R}^n$, and $H \in \mathbf{R}^n$.

Note that too many constraints may make the optimization problem infeasible. If that happens, we can reduce the constraints in $\tilde{G}$ to expand its null space. Another thing is that there might be different solutions for different support size n, so it is interesting in the future to determine under which condition the solution is unique for some n and which distribution is optimal among different n's.

4.5 Pointwise deconvolution

New signal comes in continuously at any time. So if we want to get the correct output sequence, we have to eliminate the previous signal effects, through which we may be able to do the point-wise deconvolution. But in practice, this is not possible, since what we can get are always part of the convolution results. For example, given $F = [1, 0.8, 0.6, 0.4, 0.2]$, we have a glucose sequence G as $G = [9, 8, 7, 6, 5, 4, 3, 2, 1]$ and the corresponding H measured at points when $G = 5, 4, 3, 2, 1$ as $H = [19, 16, 13, 10, 7]$. However, the true convolution results of F and G are $\hat{H} = [9, 15.2, 18.8, 20, 19, 16, 13, 10, 7, 4, 2, 0.8, 0.2]$ where we only have part of the results in practice, so the pointwise deconvolution is not possible. Approximation techniques from Fourier analysis lie in the future work.

5 Experiments

In practice, we can only measure the glucose concentration G up to some finite samples and sometimes the sample length cannot cover the support size of the density distribution. Thus in order to perform the simulations, we need to suppose the $G(t)$ is circulant, i.e., we extend the measurement sequence periodically. The simulation sequence is collected from the PhysioNet dataset by Goldberger et al. [2]. To simplify the process, we ignore the practical interval between measurements and take the interval as 1 day, in other words, we take $\Delta t = 1$.

Here is a sample used in the experiments. $G[n] = [135, 140, 169, 215, 224, 201, 265, 252, 332, 325, 296, 240, 285, 294, 276, 273, 296, 286, 349]$, where the last entry is the latest measurement. The HbA1c is computed as 10.5% by equation 1 from the mean of glucose concentration $\mathbb{E}[G] \approx 255.42 \text{mg/dL}$.

5.1 Uniform distribution

The experiments in this section explore 1) the possibility of computing the parameter M_{RBC} and 2) the importance of the bias term in equation 4.

Table 1: Results of computed M_{RBC} for uniform distribution

id	Average Glucose (mg/dL)	HbA1c (%)	M_{RBC} w/ bias (days)	M_{RBC} w/o bias (days)
1	255.42	10.5	49	58
2	174.29	7.7	51	65
3	133.56	6.3	49	67
4	128.86	6.1	48	65
5	105.86	5.3	51	74
6	108.24	5.4	50	72
7	108.05	5.4	53	75
8	110.40	5.5	52	74
9	120.05	5.8	52	73
10	225.11	9.5	48	59

Notice that in equation 5, it is hard to represent the parameter M_{RBC} by HbA1c and $G(n)$ analytically. So as an alternative, we can enumerate M_{RBC} and compute the corresponding $\mathrm{HbA1c}_{M_{\mathrm{RBC}}}$ and then select M_{RBC} that gives the closed value to HbA1c, i.e.,

$$\hat{M_{\mathrm{RBC}}} = \arg\min_{M_{\mathrm{RBC}}} \left(\mathrm{HbA1c} - \mathrm{HbA1c}_{M_{\mathrm{RBC}}}\right)^2.$$

We take 10 sequences of glucose concentration and compute the corresponding average glucose, HbA1c, M_{RBC} with and without the bias term. The results are shown in Table 1. From the table, it can be seen that M_{RBC} with bias is much more consistent for patients both with diabetes or non-diabetes, and the standard deviation is 1.77 days. On the other hand, for M_{RBC} without bias term, the standard deviation is 6.34 days. It shows that there is not huge difference from person to person as to age distributions, so we can compute the overall density function under the assumption of uniform distribution,

$$f_{\mathrm{RBC}_{\mathrm{age}}}(t) = \frac{1}{100.6}, \qquad 0 \le t \le 100.6.$$

5.2 Distinguishing uniform distribution from finite exponential distribution

The simulation takes an age distribution with support of 5 days. The setting is the same as in the previous simulation unless specified. Here the pdf of the uniform distribution is

$$f^u_{\mathrm{RBC}_{\mathrm{age}}}(x) = \frac{1}{5}, \quad x \in [0, 5]$$

and the pdf of the finite exponential distribution is

$$f^e_{\mathrm{RBC}_{\mathrm{age}}}(x) = \frac{e^{-x}}{1 - e^{-5}}, \quad x \in [0, 5]$$

The glucose concentration is increased from 100 mg/dL to 150 mg/dL and kept constant for step-like signal while decreased to 100 mg/dL after 1 day for pulse-like signal. The results are plotted in Figure 2.

For the pulse-like signal, the left one in Figure 2 shows that the pseudo[1] HbA1c increases at the pulse and then decreases linearly for uniform distribution and exponentially for finite exponential distribution. On the other hand, for the step-like signal, the day-by-day HbA1c increment difference is plotted in the right figure. It can be seen that the increment difference of the uniform distribution is decreasing linearly but the increment difference of the finite exponential distribution is decreasing exponentially.

Both these two kinds of signals can be attained easily in clinical environment and the experiment demonstrates that they can be used to distinguish distributions with different shapes.

---

[1] The reason why we call it "pesudo" is that it equals to the true HbA1c up to a constant but that is enough to illustrate the effects.
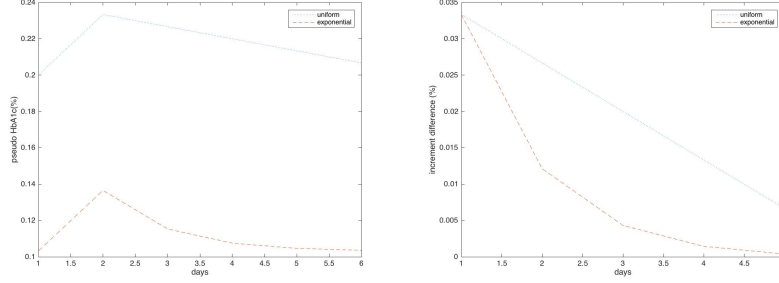
Figure 2: The left figure is for pulse-like signal and the right figure is for step-like signal.

5.3 Linear programming optimization

As in section 4.4, we construct the problem as equivalent to solving a constrained linear system, and we can solve it readily by current convex optimization softwares. For this experiment, we use CVX from Grant and Boyd [4]. We can try many different support size n and plot them for comparisons.

In the experiments, if we specify too many constraints in matrix $\tilde{G}$, the problem my be infeasible. In an extreme case, we specify only one constraint in $\tilde{G}$ to see what happens.

It turns out that if we use the same sequence as in the previous section, we still get an infeasible problem. But if we set the sequence as randomly generated, it is feasible. All the points are generated by the command `G=100+10*randn(2*n,1)`.

The cdf for $n = 60, 80, 100, 120, 140$ is plotted in Figure 3. As n increases, much density moves from old cells to young cells. It is interesting to note that the approximatly uniform distribution when $n = 100$ corresponds to our previous experiments in section 5.1, where we dictate a uniform distribution and also get $n \approx 100$.
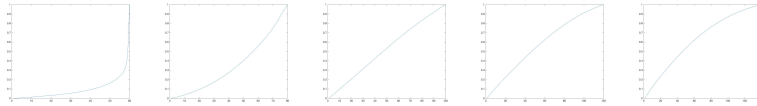

Figure 3: From left to right are the cdfs for $n = 60, 80, 100, 120, 140$.

6 Future work

1) It is obvious that the assumption of finite exponential distribution shall be investigated and experimented. 2) And the correction of the specification of the linear programming optimization problem should be further tested. It might be interesting to look into the conditions under which the solution is unique for some n and which distribution is optimal among different n's. 3) As pointed out in the previous section, the pointwise deconvolution is not possible even though we model the problem as a convolution equation. The approximation method in the frequency domain by using Fourier analysis is worth investigating.

7 Conclusion

In this paper we define the connections between HbA1c, $G(t)$, and $f_{\mathrm{RBC_{age}}}(t)$ as a convolution operation. Necessary assumptions for feasible inference of $f_{\mathrm{RBC_{age}}}$ are provided. Experiments show that our proposed assumptions are consistent at least if the cell age distribution is uniform. Other theoretical aspects involving how to distinguish the uniform distribution and the finite exponential distribution, how to specify a linear programming problem for inferring a generic distribution, and the possibility of applying the pointwise deconvolution are also investigated.

References

[1] AH Berg and DB Sacks. Haemoglobin a1c analysis in the management of patients with diabetes: from chaos to harmony. *Journal of clinical pathology*, 61(9):983–987, 2008.

[2] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet. *Circulation*, 101(23):e215–e220, 2000.

[3] David E Goldstein. Is glycosylated hemoglobin clinically useful?, 1984.

[4] Michael Grant and Stephen Boyd. Cvx: Matlab software for disciplined convex programming.

[5] Richard Kahn, Jocelyn Hicks, Mathias Muller, Mauro Panteghini, Garry John, Larry Deeb, John Buse, David M Nathan, Ele Ferrannini, Robert Heine, et al. Consensus statement on the worldwide standardization of the hemoglobin a1c measurement: the american diabetes association, european association for the study of diabetes, international federation of clinical chemistry and laboratory medicine, and the international diabetes federation. *Diabetes Care*, 30 (9):2399–2400, 2007.

[6] David M Nathan, Daniel E Singer, Katherine Hurxthal, and John D Goodson. The clinical information value of the glycosylated hemoglobin assay. *New England Journal of Medicine*, 310 (6):341–346, 1984.

[7] David M Nathan, Judith Kuenen, Rikke Borg, Hui Zheng, David Schoenfeld, and Robert J Heine. Translating the a1c assay into estimated average glucose values. *Diabetes care*, 31(8):1473–1478, 2008.

[8] DM Nathan, H Turgeon, and S Regan. Relationship between glycated haemoglobin levels and mean glucose levels over time. *Diabetologia*, 50(11):2239–2244, 2007.