Statistics and Probability

- Read Chapter 3.
- What is error?
- Two kinds: Random & Systematic
- Systematic error is <u>systematic</u>: it happens to the same degree <u>each time</u> you make a measurement.
- Multiple measurements will average out (reduce) random error, but <u>won’t</u> affect systematic error!

Statistics and Probability

- If you do many measurements, the error in the average will be <u>completely</u> dominated by systematic error, but this is very difficult to estimate!
- Be wary of error estimates – always look to see how it is calculated.
- As an engineer, a detailed understanding of statistics and error is <u>critical</u>: you <u>must</u> master this material!

Statistics and Probability

- Now for some elementary statistics:
- The most important <u>probability distribution</u> is the Gaussian or normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$

- f(x) is the probability density.
- f(x) = probability of measurement being in the interval [x, x+dx]/dx

Statistics and Probability

- We can define the cumulative probability:

$$P(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$

- This is the probability that a measurement is less than some value x.
- The normal distribution is characterized by 2 quantities:
  - μ = mean of distribution
  - σ = population standard deviation

Statistics and Probability

- About 68% of observations lie within 1σ of μ, 95% within 2σ, and 99% within 3σ.
- In working with statistics, we define an <u>expectation value</u>

  E(x) = what you expect to get if you do a measurement many times and average it together.

The **expected value** or **expectation** of a random variable x that has a continuous probability distribution $f(x)$ is equal to the mean μ of the distribution and is defined as

$$E(x) = \mu = \int_{-\infty}^{\infty} x f(x) dx. \tag{3.24}$$

Expectation value

- What is the expectation value of a random variable “x” that follows a Normal distribution?

$$E(x) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}\, dx$$

$$= \int_{-\infty}^{\infty} \frac{x-\mu+\mu}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}\, dx$$

Integration by parts

$$= \int_{-\infty}^{\infty} \frac{(x-\mu)}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}\, dx + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}\, dx.$$

Substituting $z = x-\mu$ and $\int_{-\infty}^{\infty} f(x)dx = 1$ (Equation (3.21)), we get

$$E(x) = \int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}\sigma} e^{-z^2/2\sigma^2}\, dz + \mu \cdot 1.$$

Expectation value

$$E(x) = \int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}\sigma} e^{-z^2/2\sigma^2}\ dz + \mu \cdot 1.$$

- Odd function * even function = odd function. Odd function integrated over an even interval =0.
- Thus, **E(x) = μ**. We just proved that μ represents the mean of the normal distribution.
- What about the mean of n observations?

$$E(\bar{x}) = E\left(\frac{1}{n}\sum_{x=1}^{n} x_i\right) = \frac{1}{n}\sum_{x=1}^{n} E(x_i) = \frac{1}{n}\sum_{x=1}^{n} \mu = \mu.$$

Expectation value

For any continuous probability distribution $f(x)$, the **variance** is the expected value of $(x - \mu)^2$ and is defined as

$$\sigma^2 = E((x-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx. \tag{3.25}$$

The variance of a normal distribution $N(\mu, \sigma)$ is calculated by substituting Equation (3.23) into Equation (3.25):

$$\sigma^2 = E((x-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}\, dx.$$

Let $z = (x - \mu)$. Then

$$\sigma^2 = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} z^2 \cdot e^{-z^2/2\sigma^2}\, dz$$

The function $z^2 e^{-z^2/2\sigma^2}$ is an even function. Therefore the interval of integration can be halved:

Expectation value

$$\frac{2}{\sqrt{2\pi}\sigma}\int_0^{\infty} z^2 \cdot e^{-z^2/2\sigma^2}\ dz.$$

Let $y = z^2/2$. Then $dy = z\ dz$. The integral becomes

$$\frac{2}{\sqrt{2\pi}\sigma}\int_0^{\infty} \sqrt{2y} \cdot e^{-y/\sigma^2}\ dy.$$

Integrating by parts $\left(\int uv\ dy = u\int v\ dy - \int\left[u' \int v\ dy\right]dy\right)$, we obtain

$$E((x-\mu)^2) = \frac{2}{\sqrt{2\pi}\sigma}\left[\sqrt{2y}\cdot e^{-y/\sigma^2}\cdot\left(-\sigma^2\right)\Big|_0^{\infty} - \int_0^{\infty}\frac{1}{\sqrt{2y}}\cdot\left(-\sigma^2\right)\cdot e^{-y/\sigma^2}\ dy\right].$$

The first term on the right-hand side is zero. In the second term we substitute z back into the equation to get

$$E\left((x-\mu)^2\right) = 0 + 2\sigma^2\int_0^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-z^2/2\sigma^2}\ dz = 2\sigma^2\cdot\frac{1}{2} = \sigma^2.$$

Thus, the variance of a normal distribution is σ^2, and the standard deviation is σ.

The variance of the distribution of sample means is defined as

$$s_{\bar{x}}^2 = E\left((\bar{x} - \mu)^2\right).$$

Now,

$$E\left((\bar{x} - \mu)^2\right) = E\left(\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \mu\right)^2\right) = \frac{1}{n^2} E\left(\left(\sum_{i=1}^{n} x_i - n\mu\right)^2\right)$$

$$= \frac{1}{n^2} E\left(\left(\sum_{i=1}^{n}(x_i - \mu)\right)^2\right).$$

The squared term can be expanded using the following rule:

$$(a + b + c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc.$$

So

$$E\left((\bar{x} - \mu)^2\right) = \frac{1}{n^2} E\left(\sum_{i=1}^{n}(x_i - \mu)^2 + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}(x_i - \mu)(x_j - \mu)\right).$$

The double summation in the second term will count each $(x_i - \mu)(x_j - \mu)$ term twice. The order in which the expectations and the summations are carried out is now switched:

$$E\left((\bar{x} - \mu)^2\right) = \frac{1}{n^2}\sum_{i=1}^{n} E\left((x_i - \mu)^2\right) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} E((x_i - \mu)(x_j - \mu)).$$

Expectation value

Let's investigate the double summation term:

$$E\big((x_i-\mu)(x_j-\mu)\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_i-\mu)(x_j-\mu)f(x_i)f(x_j)\,dx_i\,dx_j.$$

If x_i and x_j are independent measurements (i.e. the chance of observing x_i is not influenced by the observation of x_j), then

$$E\big((x_i-\mu)(x_j-\mu)\big) = \int_{-\infty}^{\infty}(x_i-\mu)f(x_i)\,dx_i\int_{-\infty}^{\infty}(x_j-\mu)f(x_j)\,dx_j = 0 \text{ for } i \neq j.$$

Therefore,

$$E\Big((\bar{x}-\mu)^2\Big) = \frac{1}{n^2}\left(\sum_{i=1}^{n}E\Big((x_i-\mu)^2\Big)\right)+0.$$

By definition, $E\Big((x_i-\mu)^2\Big)=\sigma^2$, and

$$s_{\bar{x}}^2 = E\Big((\bar{x}-\mu)^2\Big) = \frac{\sigma^2}{n}. \qquad (3.27)$$

SEM: Standard Error of the Mean

The standard deviation of the distribution of sample means is

$$s_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{3.28}$$

and is called the **standard error of the mean** or **SEM**.

- That is why multiple measurements reduce the random error!

make use of the following identity:

$$\sum_{i=1}^{n}(x_i-\mu)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2.$$

Equation 3.29 can be easily proven by expanding the terms. For example,

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})^2 &= \sum_{i=1}^{n}(x_i^2-2x_i\bar{x}+\bar{x}^2)\\ &= \sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+n\bar{x}^2\\ &= \sum_{i=1}^{n}x_i^2-n\bar{x}^2.\end{aligned}$$

Therefore

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}(x_i^2-\bar{x}^2).$$

Sample variance

Substituting Equation (3.29) into Equation (3.3), we get

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i - \mu)^2 - n(\bar{x} - \mu)^2}{n-1}.$$

Then, the expected variance of the sample is given by

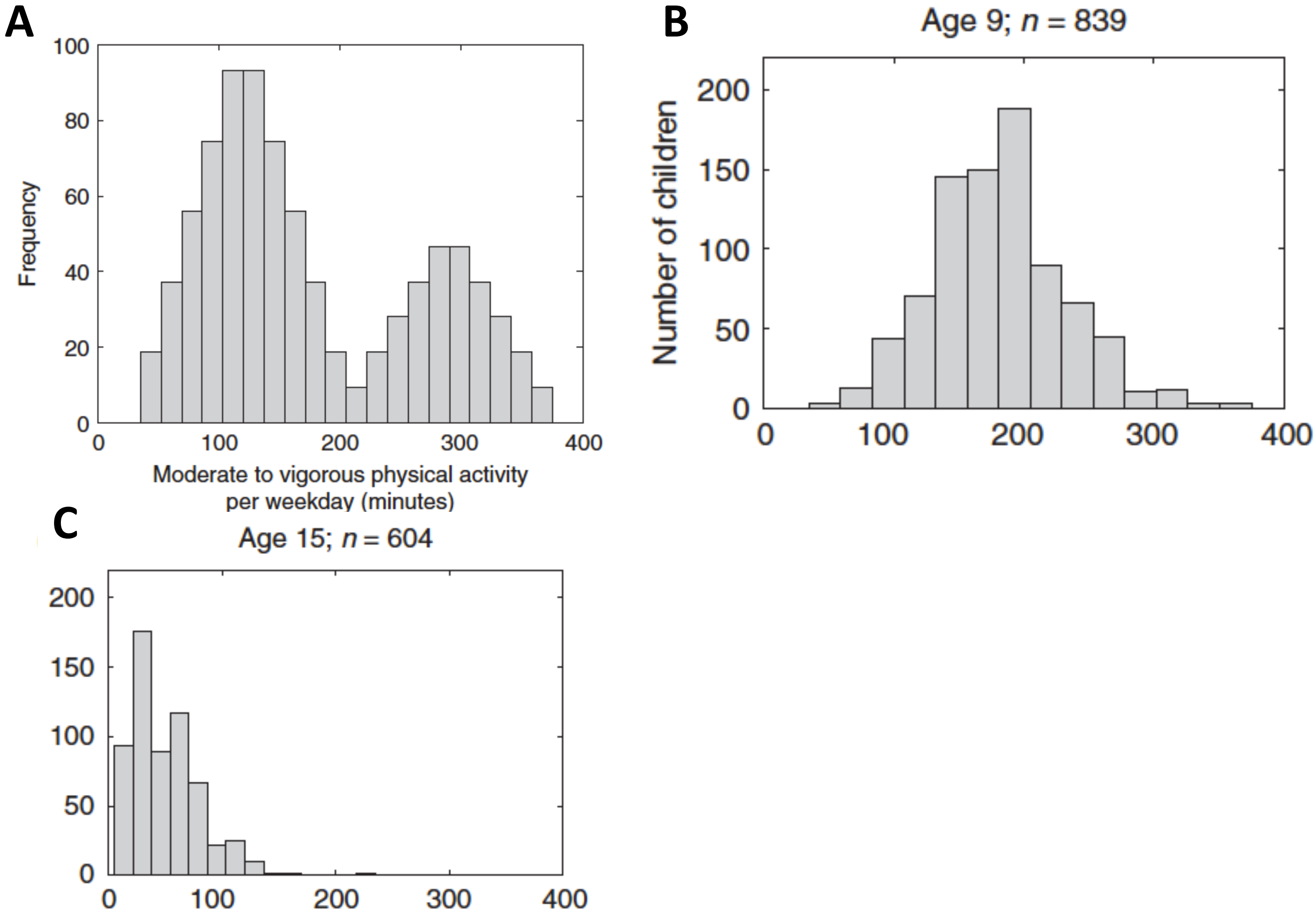
$$E\left(s^2\right) = \frac{\sum_{i=1}^{n} E\left((x_i - \mu)^2\right) - nE\left((\bar{x} - \mu)^2\right)}{n-1}.$$

From Equations (3.25) and (3.27),

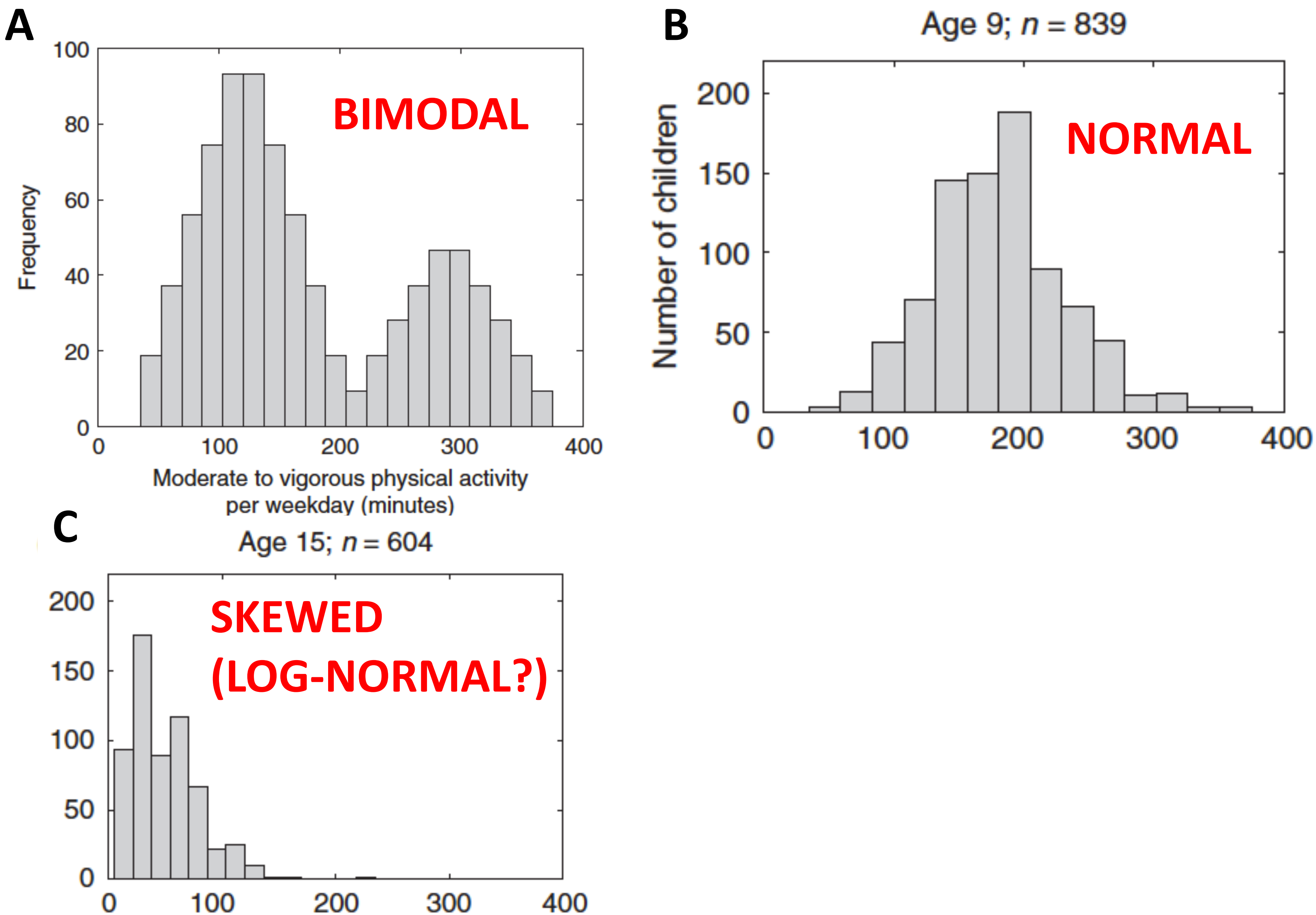
$$E\left(s^2\right) = \frac{\sum_{i=1}^{n} \sigma^2 - n\frac{\sigma^2}{n}}{n-1} = \frac{n\sigma^2 - \sigma^2}{n-1} = \sigma^2.$$

Thus, s^2 is an unbiased estimate for σ^2!

Q1: Which of these distributions is approximately “Normal”?



Q1: Which of these distributions is approximately "Normal"?



Q2: Which of the normal density curves has the greatest area under the curve?

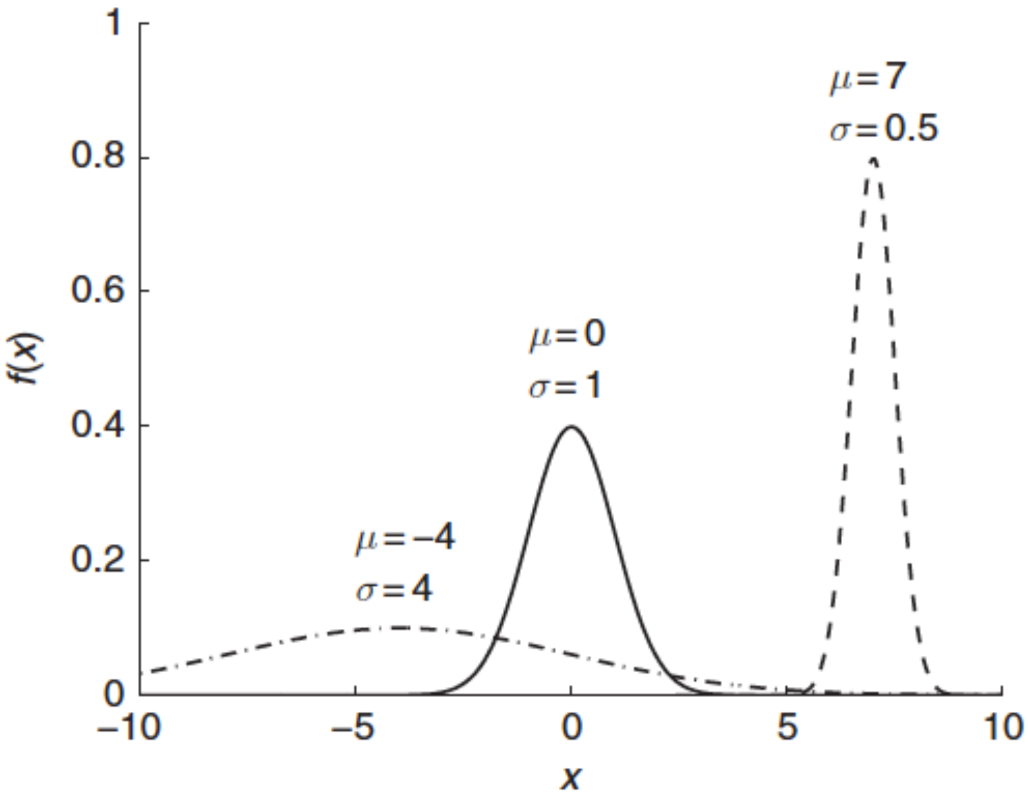
A. A

B. B

C. C

D. A = B = C

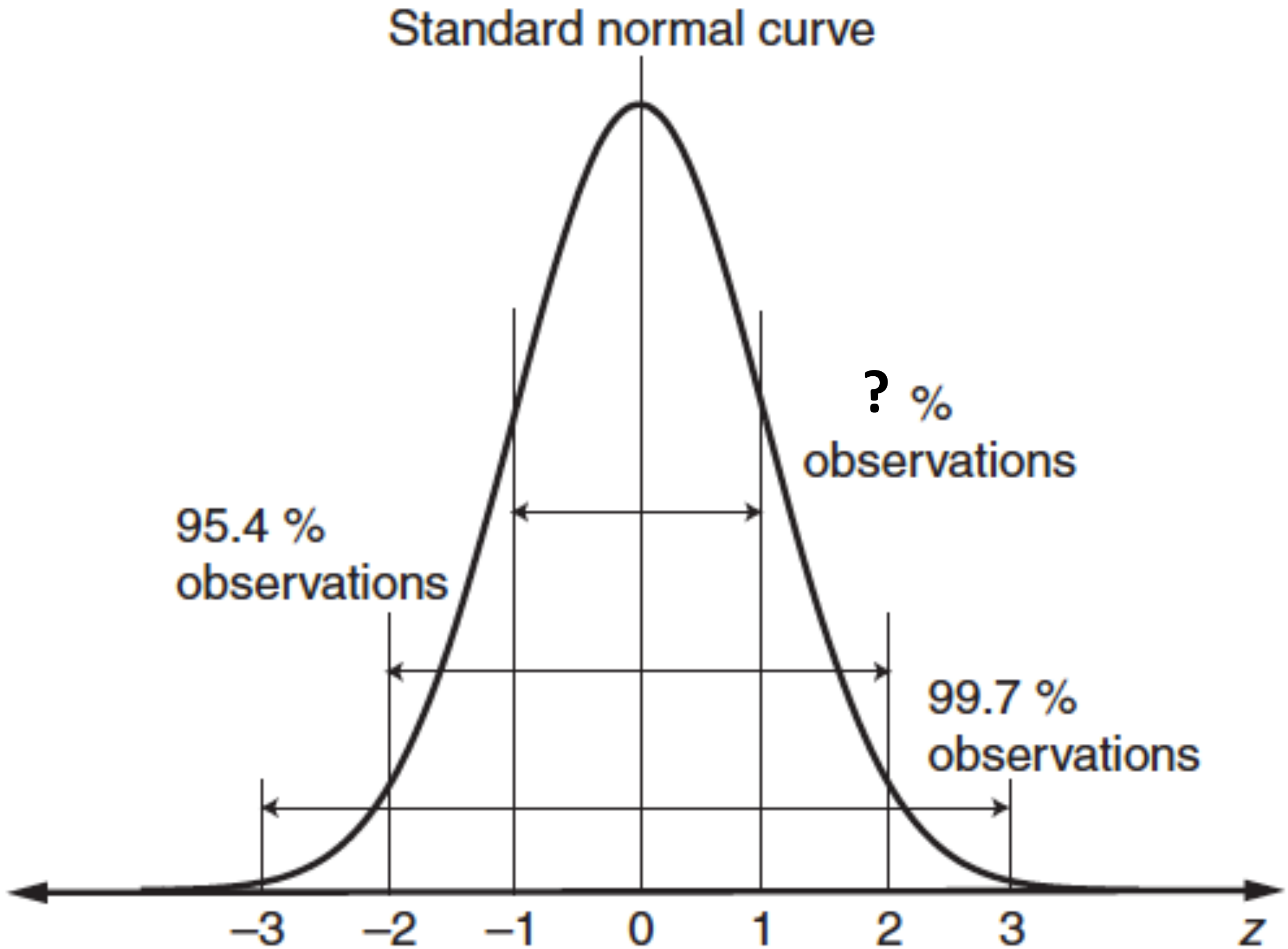
Three normal density curves, which include the standard normal probability distribution ($\mu = 0$, $\sigma = 1$).



Q3: In a normal distribution, WHAT % of observations lie within ±1 σ of the mean value?

A. 50%
B. 68%
C. 75%
D. 83%

Areas under the standard normal curve.



Q4: The standard error of the mean (SEM) is the standard deviation of sample means. If the calculated standard deviation of the sample is σ, then SEM is equal to:

A. $$SEM = \sigma^2 / n$$

B. $$SEM = \sigma / \sqrt{n}$$

C. $$SEM = \sigma / n$$

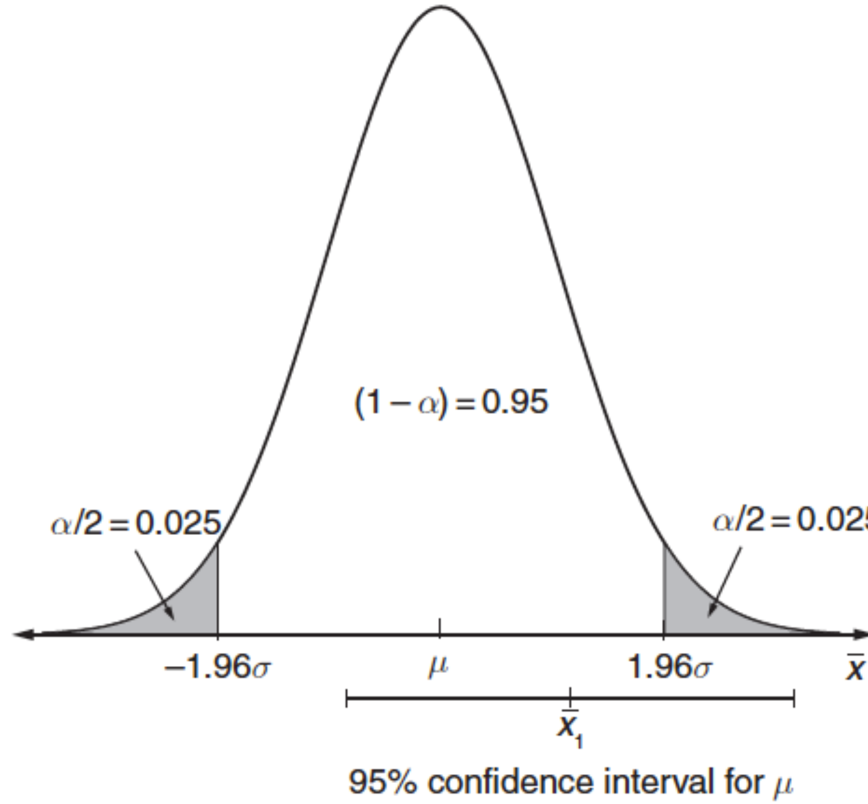
D. $$SEM = n / \sigma^2$$

Q4: The standard error of the mean (SEM) is th
If the calculated standard deviation of the sam

A. $$SEM = \sigma^2 / n$$

B. $$\boxed{SEM = \sigma / \sqrt{n}}$$

C. $$SEM = \sigma / n$$



If the interval $\mu \pm 1.96\ \sigma/\sqrt{n}$ contains 95% of the sample means, then 95% of the intervals $\bar{x} \pm 1.96\ \sigma/\sqrt{n}$ must contain the population mean.

Langmuir **2007**, *23*, 6321−6328

Influence of Brownian Motion on Blood Platelet Flow Behavior and Adhesive Dynamics near a Planar Wall

Nipa A. Mody and Michael R. King*

Departments of Chemical Engineering and Biomedical Engineering, University of Rochester, Rochester, New York 14627

Received January 17, 2007. In Final Form: February 27, 2007

We used the platelet adhesive dynamics computational method to study the influence of Brownian motion of a platelet on its flow characteristics near a surface in the creeping flow regime. Two important characterizations were done in this regard: (1) quantification of the platelet's ability to contact the surface by virtue of the Brownian forces and torques acting on it, and (2) determination of the relative importance of Brownian motion in promoting surface encounters in the presence of shear flow. We determined the Peclet number for a platelet undergoing Brownian motion in shear flow, which could be expressed as a simple linear function of height of the platelet centroid, H from the surface $Pe\ (\text{platelet}) = \dot{\gamma} \cdot (1.56H + 0.66)$ for $H > 0.3\ \mu\text{m}$. Our results demonstrate that at timescales relevant to shear flow in blood Brownian motion plays an insignificant role in influencing platelet motion or creating further opportunities for platelet−surface contact. The platelet Peclet number at shear rates $>100\ \text{s}^{-1}$ is large enough (>200) to neglect platelet Brownian motion in computational modeling of flow in arteries and arterioles for most practical purposes even at very close distances from the surface. We also conducted adhesive dynamics simulations to determine the effects of platelet Brownian motion on GPIbα-vWF-A1 single-bond dissociation dynamics. Brownian motion was found to have little effect on bond lifetime and caused minimal bond stressing as bond rupture forces were calculated to be less than 0.005 pN. We conclude from our results that, for the case of platelet-shaped cells, Brownian motion is not expected to play an important role in influencing flow characteristics, platelet−surface contact frequency, and dissociative binding phenomena under flow at physiological shear rates ($>50\ \text{s}^{-1}$).

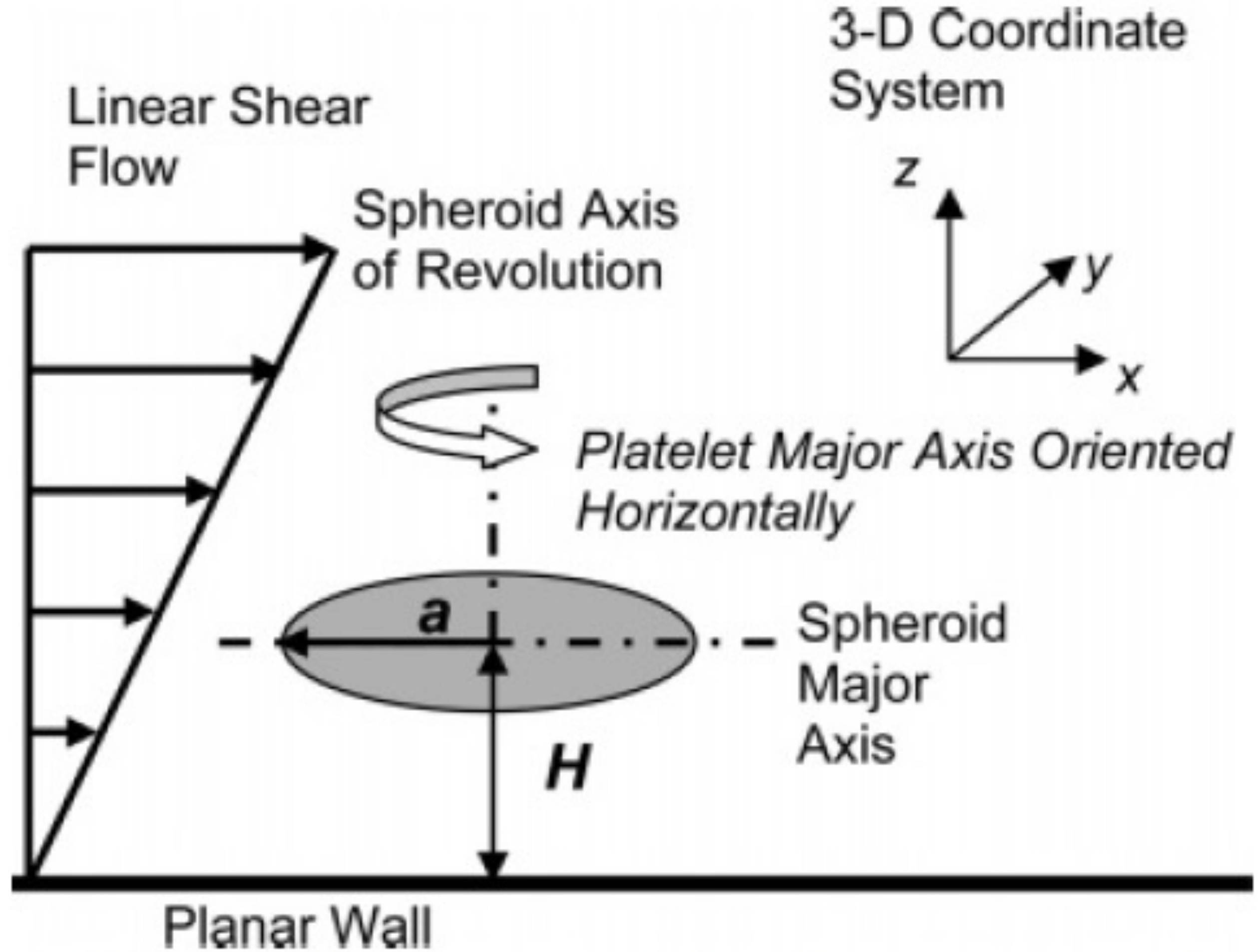


Figure 1. Schematic diagram of the flow geometry in which a single platelet (oblate spheroid of aspect ratio = 0.25) is translating and rotating in linear shear flow near an infinite planar surface. At the instant shown in the figure, the platelet is oriented with its major axis parallel to the surface and its centroid is located at a distance H from the surface.

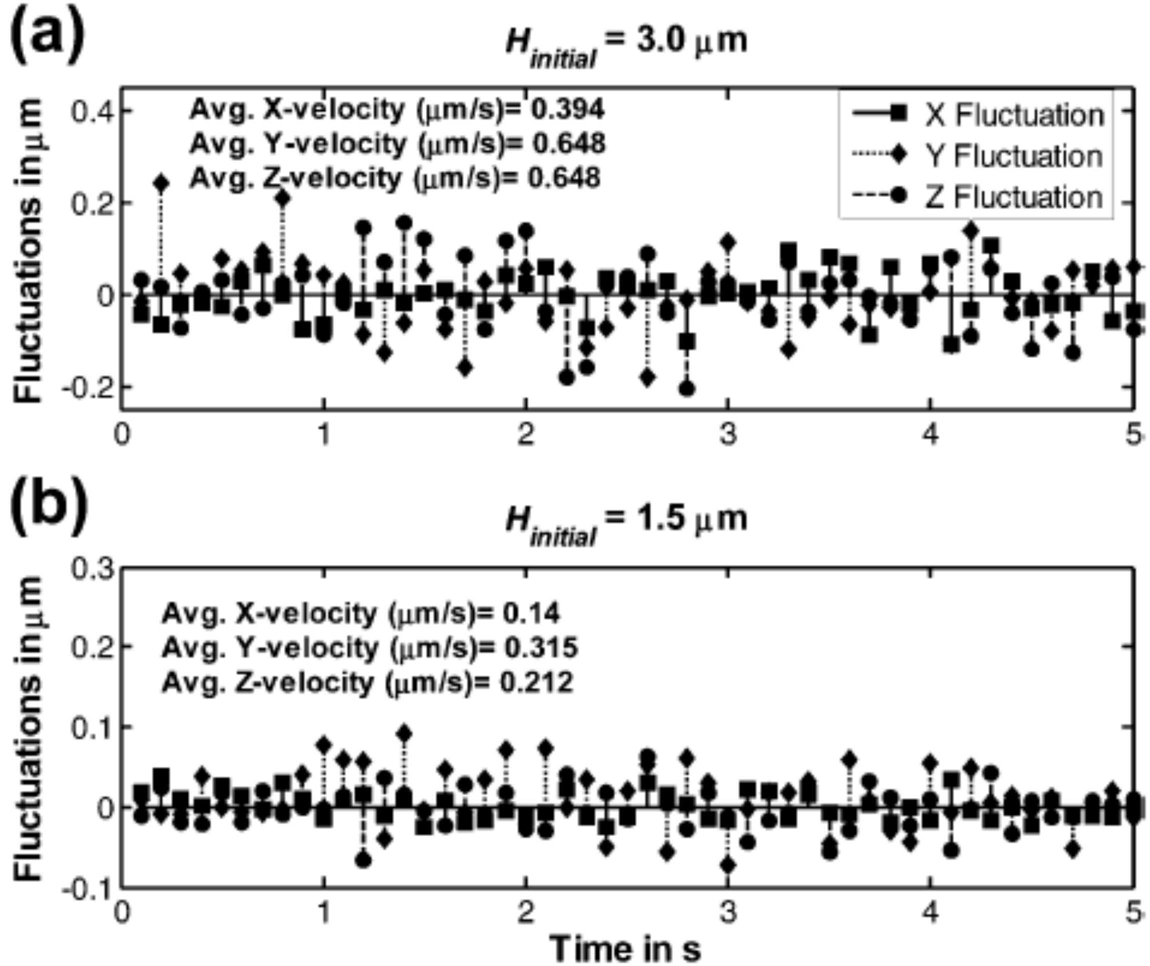


Figure 2. Plot of the distances translated in the x, y, and z directions in each time step of 0.1 s by a platelet undergoing Brownian motion near a surface in a quiescent fluid of viscosity 1 cP in the Stokes regime of flow. Fluctuations in platelet position with respect to its previous position are shown for a period of 5 s for (a) an initial platelet centroid height of 3.0 μm and (b) an initial platelet centroid height of 1.5 μm. The average velocities in each direction over the 5 s interval are listed in each plot.

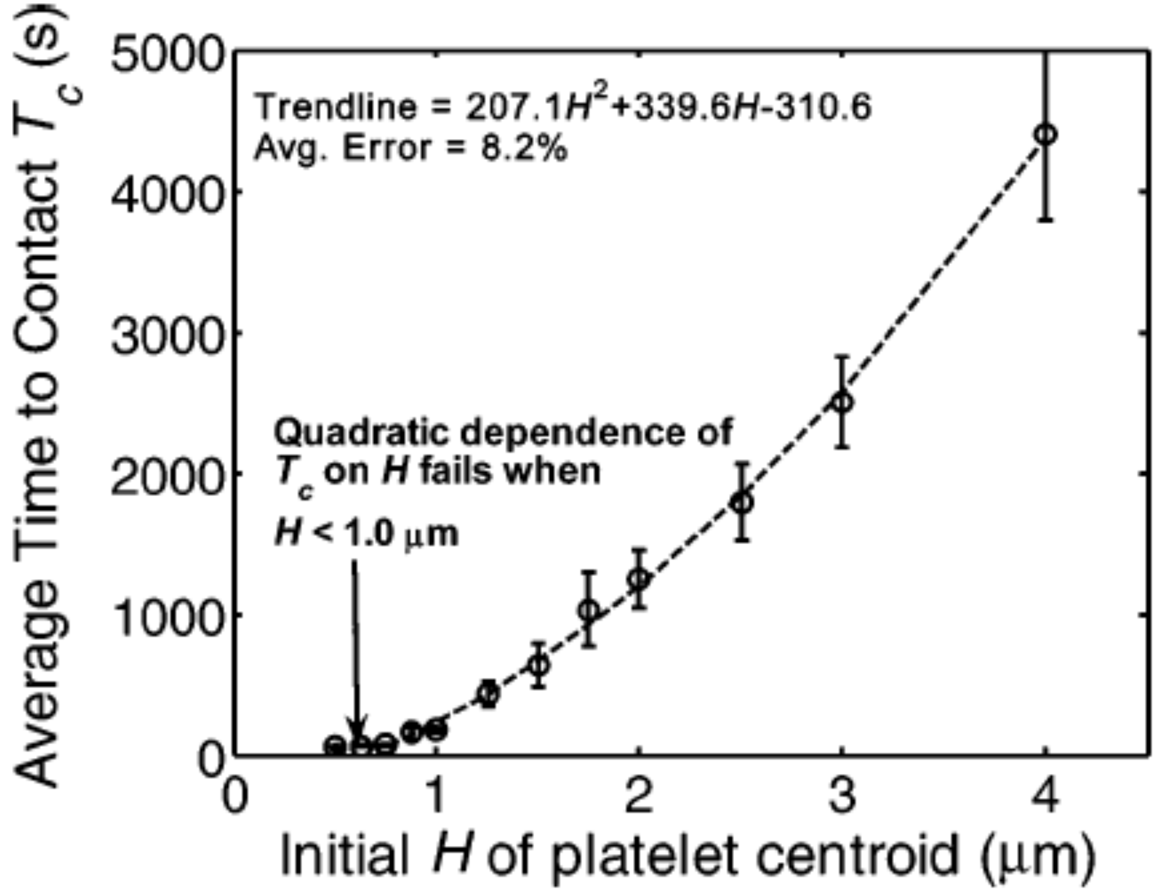


Figure 3. Plot of the average time taken T_c for a platelet at an initial height H to contact the surface when undergoing translational and rotational Brownian motion in a quiescent fluid in the Stokes regime of flow. The dashed trendline does not include data points at initial $H = 0.5$ and $0.625\ \mu$m.

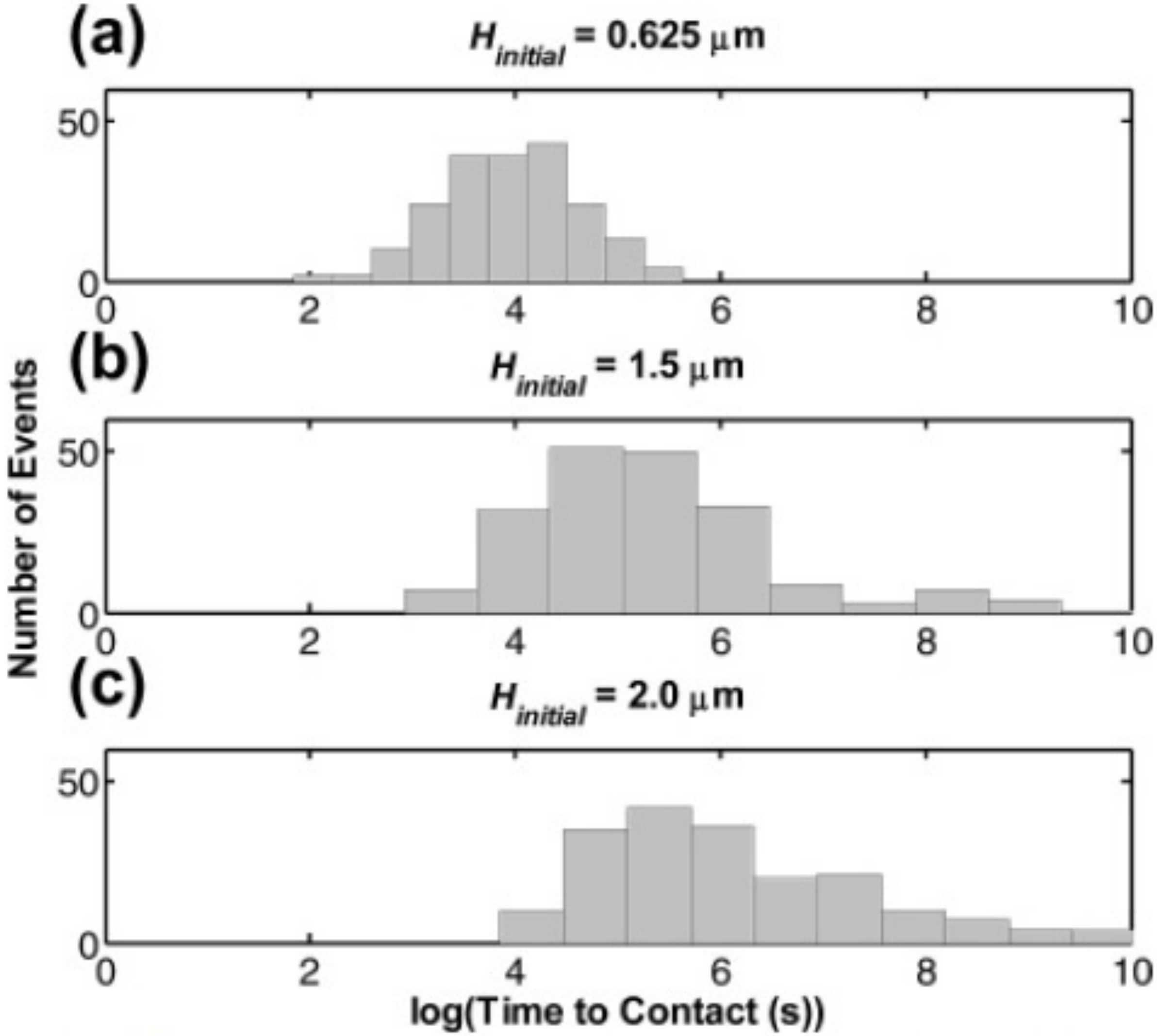


Figure 4. Histograms of the log of time taken T_c for a platelet to contact the surface when undergoing translational and rotational Brownian motion in a quiescent fluid for three different initial heights H: (a) $H_{\text{initial}} = 0.625$ μm, (b) $H_{\text{initial}} = 1.5$ μm, and (c) $H_{\text{initial}} = 2.0$ μm, representing approximately 200 cell-surface contact events for each H_{initial}.

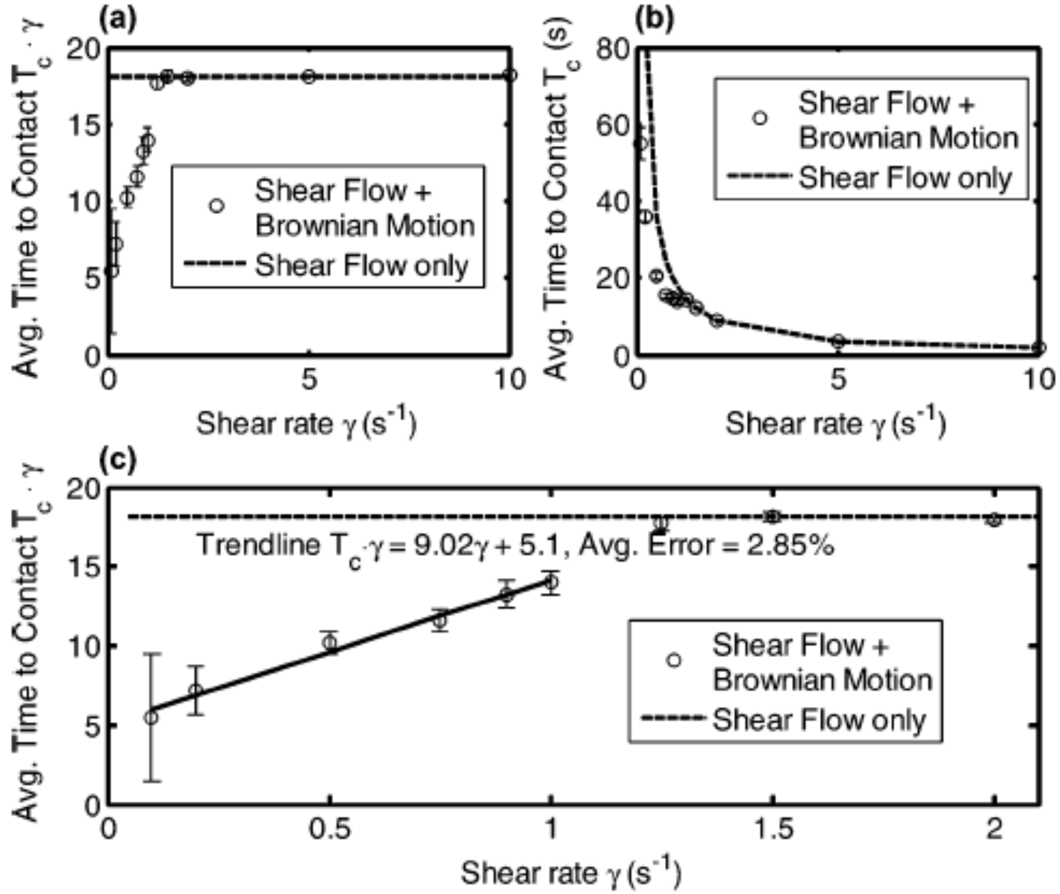


Figure 5. Plots of the time until surface contact T_c as a function of shear rate for a Brownian platelet at an initial centroid height of 0.8 μm in linear shear flow. The dashed line indicates "time until surface contact" for platelets undergoing linear shear flow only; that is, Brownian force and torque were not included in the mobility calculations. In panel a, "time to contact" the surface for a platelet is nondimensionalized with shear rate, and in panel b, "time to contact" is kept in dimensional form. Plot (c) is a magnified version of panel a with a trendline plotted to demonstrate the simple linear dependency of non-dimensionalized "time to surface contact" on shear rate when Brownian motion (diffusion) is the dominant mode of cell transport to the surface.

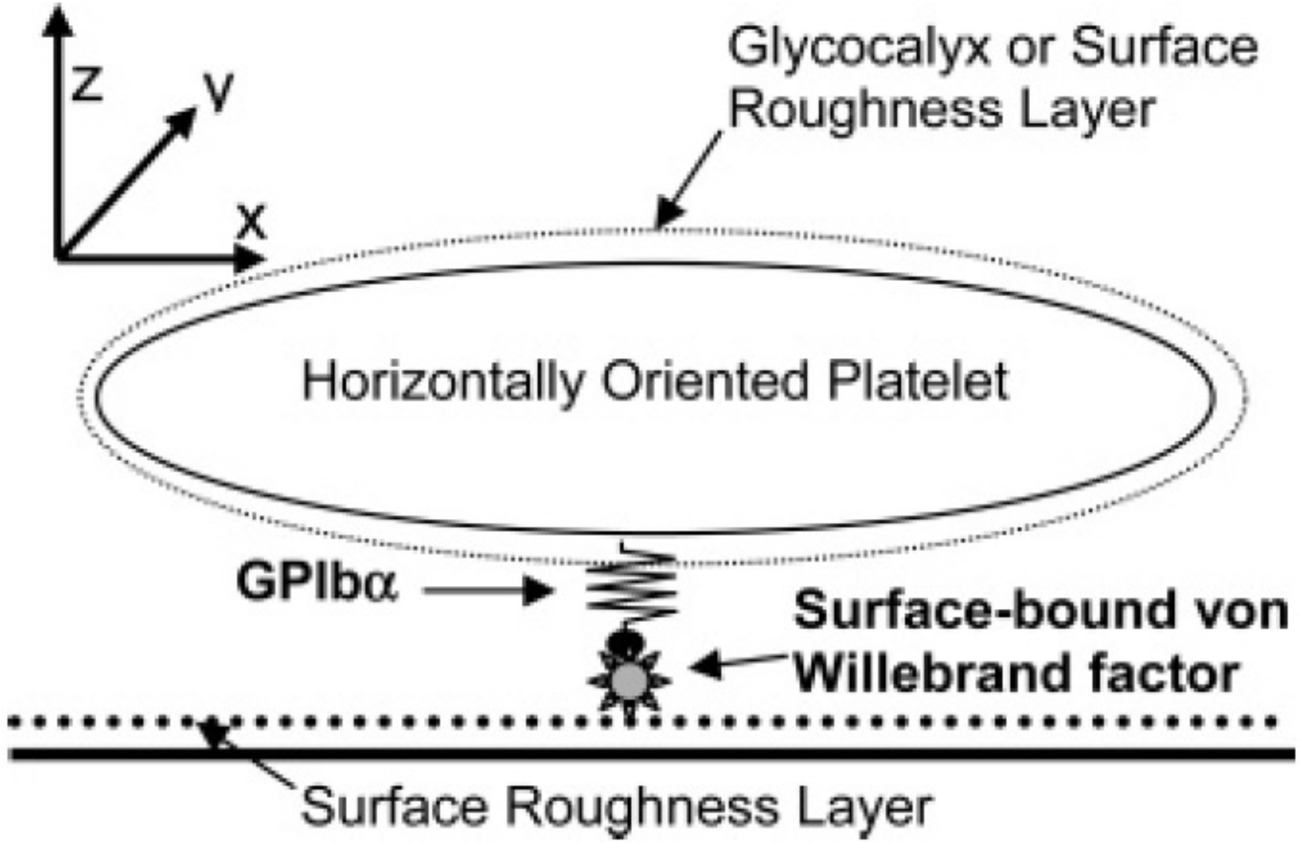


Figure 6. Schematic diagram of a horizontally oriented platelet bound to the surface via a single GPIbα-vWF bond. Both the surface and platelet are coated with a surface roughness layer of 25 nm.

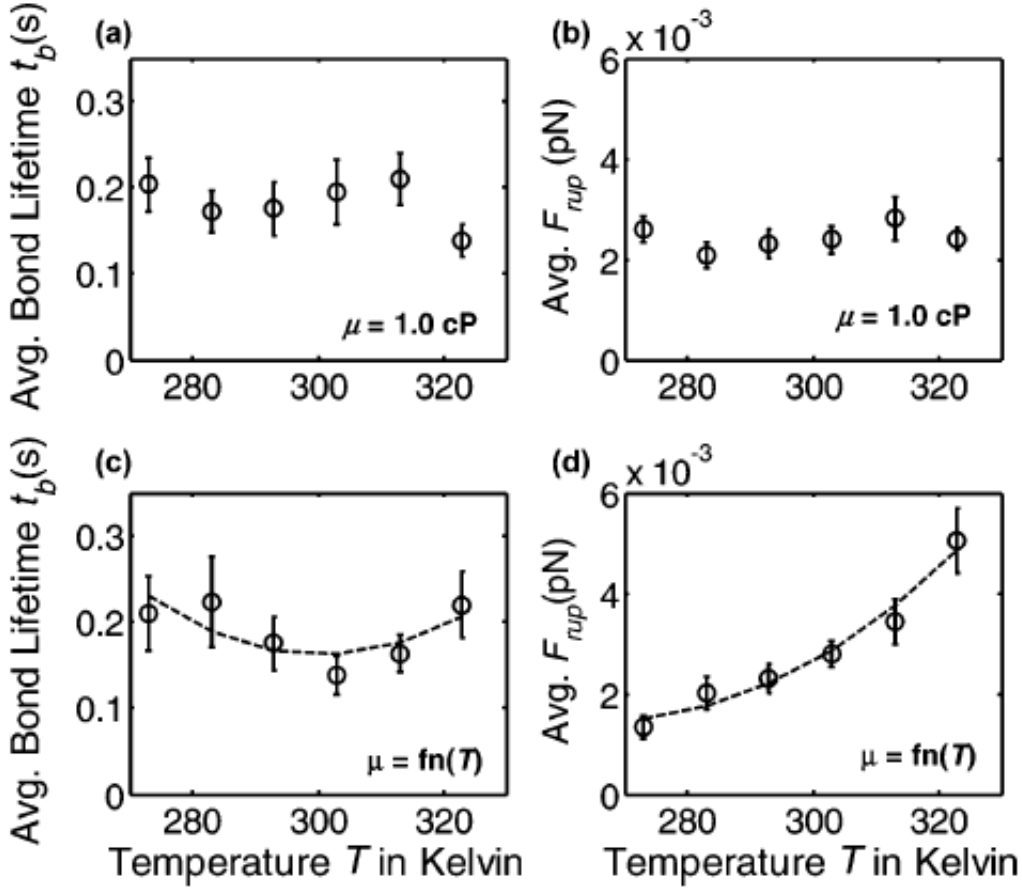


Figure 7. Plots of GPIbα-vWF-A1 bond lifetimes and bond rupture forces for a platelet undergoing Brownian motion in a quiescent fluid. (a) Bond lifetimes when the fluid medium has constant viscosity = 1.0 cP. (b) Bond rupture forces when the fluid medium has constant viscosity = 1.0 cP. For both panels a and b, 41 observations were collected for each temperature. (c) Bond lifetimes for a fluid medium with temperature-dependent viscosity. (d) Bond rupture forces for a fluid medium with temperature-dependent viscosity. For both panels c and d, 37 observations were collected for each temperature. Quadratic trendlines drawn in panels c and d are within 10% and 7.8% average error, respectively.