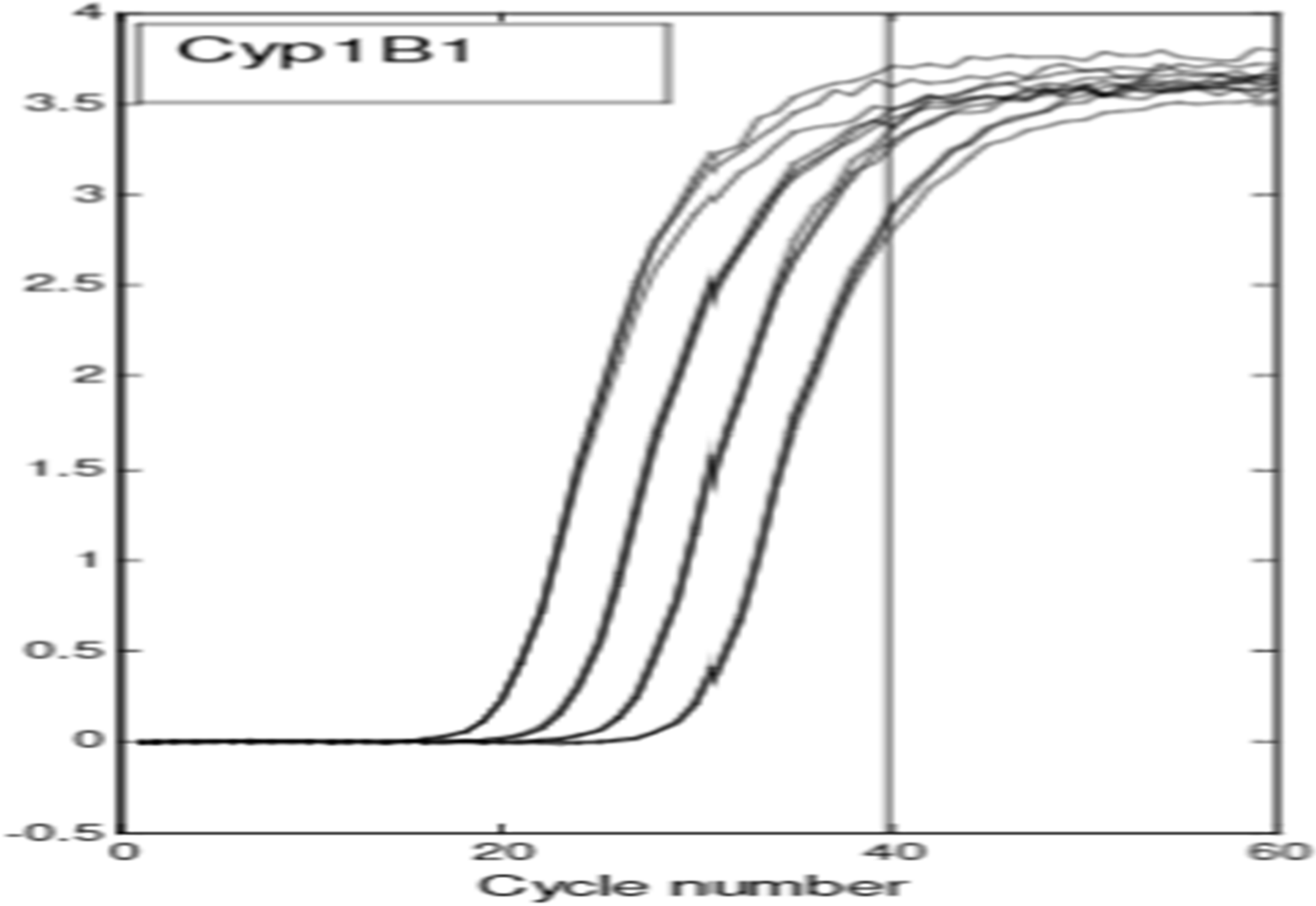
An example of the uniform distribution

Cycle threshold (Ct) value in COVID-19 infection

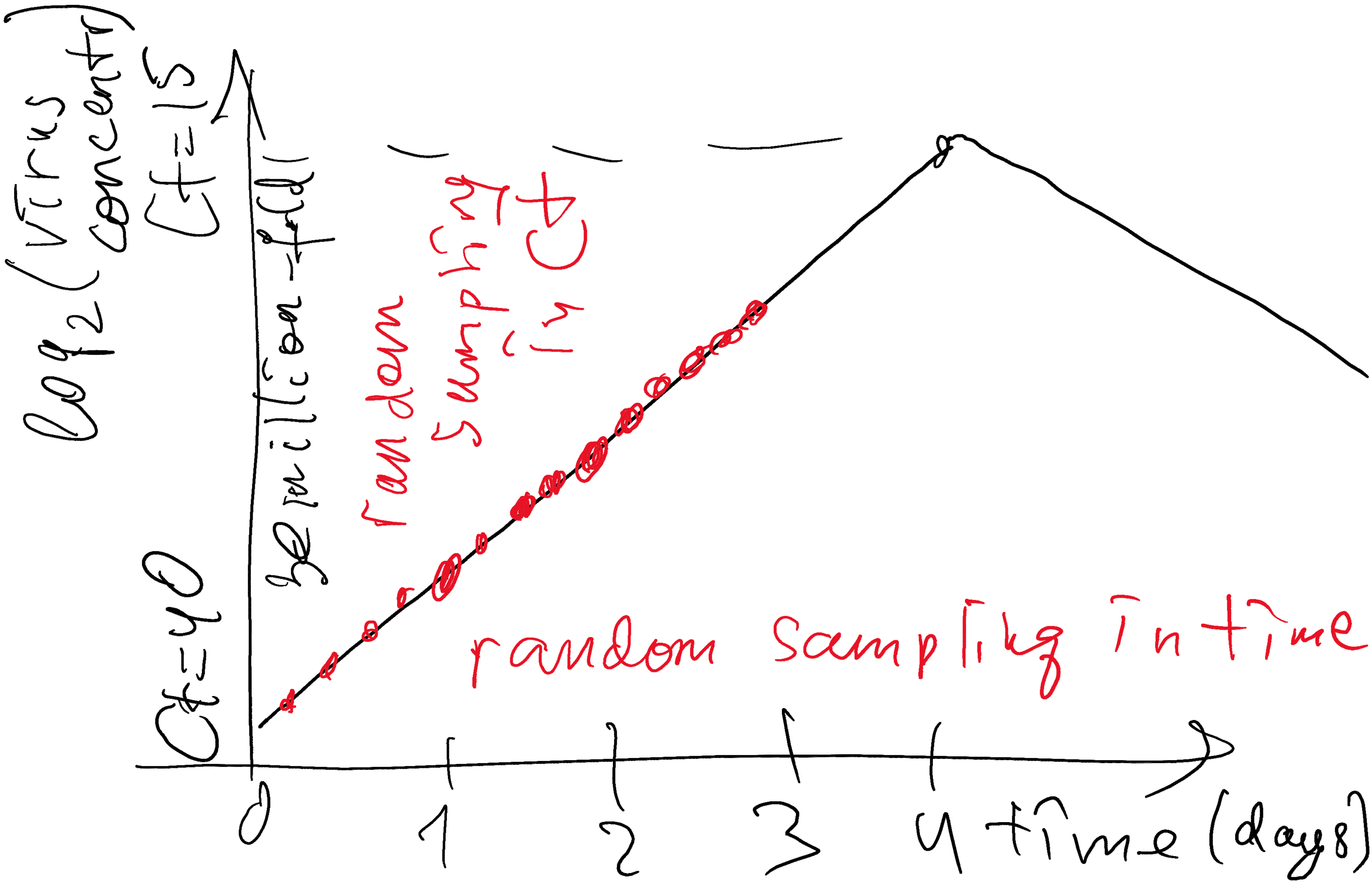
What is the Ct value of a PCR test?

Ct = const – log2(viral DNA concentration)



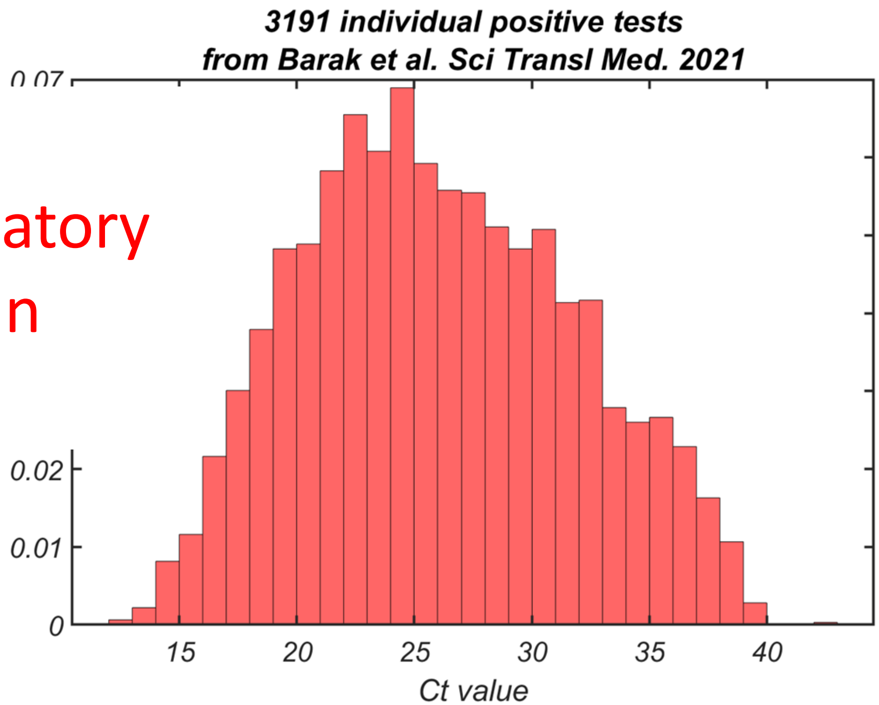
Why Ct distribution should be uniform?

Why Ct distribution should be uniform?

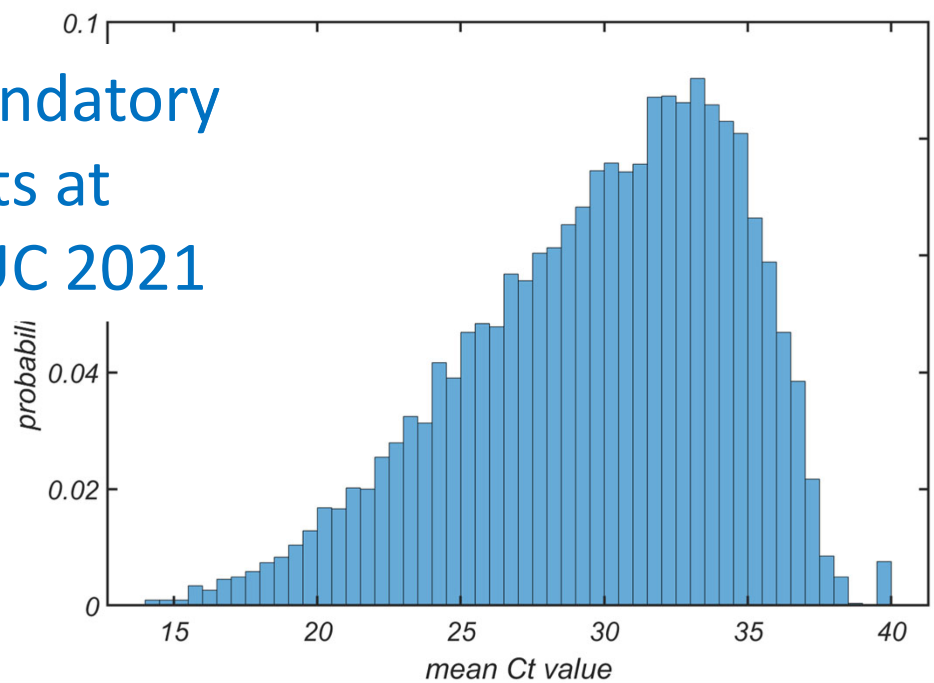


Why should we care?

Non-mandatory tests in Israel



Mandatory tests at UIUC 2021



- High Ct value means we identified the infected individual early, hopefully before transmission to others
- When testing is mandatory, and people are tested frequently – the mean Ct value is shifted towards high values

Matlab exercise: Uniform distribution

- Generate a sample of size 100,000 for uniform random variable X taking values 1,2,3,…10
- Plot the approximation to the probability mass function based on this sample
- Calculate mean and variance of this sample and compare it to infinite sample predictions: $E[X]=(a+b)/2$ and $V[X]=((a-b+1)^2-1)/12$

Matlab template: Uniform distribution

- **b=10; a=1; % b= upper bound; a= lower bound (inclusive)'**
- **Stats=100000; % sample size to generate**
- **r1=rand(Stats,1);**
- **r2=floor(??*r1)+??;**
- **mean(r2)**
- **var(r2)**
- **std(r2)**
- **[hy,hx]=hist(r2, 1:10); % hist generates histogram in bins 1,2,3...,10**
- **% hy - number of counts in each bin; hx - coordinates of bins**
- **p_f=hy./??; % normalize counts to add up to 1**
- **figure; plot(??,p_f, 'ko-'); ylim([0, max(p_f)+0.01]); % plot the PMF**

Matlab exercise: Uniform distribution

- **b=10; a=1; % b= upper bound; a= lower bound (inclusive)'**
- **Stats=100000; % sample size to generate**
- **r1=rand(Stats,1);**
- **r2=floor((b-a+1).*r1)+a;**
- **mean(r2)**
- **var(r2)**
- **std(r2)**
- **[hy,hx]=hist(r2, 1:10); % hist generates histogram in bins 1,2,3...,10**
- **% hy - number of counts in each bin; hx - coordinates of bins**
- **p_f=hy./sum(hy); % normalize counts to add up to 1**
- **figure; plot(hx,p_f, 'ko-'); ylim([0, max(p_f)+0.01]); % plot the PMF**

Bernoulli distribution

The simplest non-uniform distribution

p – probability of success (1)

1-p – probability of failure (0)

$$f(x) = P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

Jacob Bernoulli
(1654-1705)
Swiss mathematician (Basel)

- Law of large numbers
- Mathematical constant e=2.718…

Bernoulli distribution

$$f(x) = P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

$$E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) = 0(1 - p) + 1(p) = p$$

$$\text{Var}(X) = E(X^2) - (EX)^2 = [0^2(1 - p) + 1^2(p)] - p^2 = p - p^2 = p(1 - p)$$

Refresher: Binomial Coefficients

$$\binom{n}{k} = C_k^n = \frac{n!}{k!(n-k)!}\text{, called } n \text{ choose } k$$

$$\binom{10}{3} = C_3^{10} = \frac{10!}{3!7!} = \frac{10\cdot 9\cdot 8\cdot 7!}{3\cdot 2\cdot 1\cdot 7!} = 120$$

Number of ways to choose k objects out of n without replacement and where the order does not matter.

Called binomial coefficients because of the binomial formula

$$(p+q)^n = (p+q)\times(p+q)\ldots\times(p+q) = \sum_{x=0}^{n} C_x^n p^x q^{n-x}$$

Binomial Distribution

- Binomially-distributed random variable *X* equals sum (number of successes) of n independent Bernoulli trials
- The probability mass function is:

$q = 1 - p$

$$f(x) = C_x^n p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, \ldots n \qquad (3\text{-}7)$$

- Based on the binomial expansion:

$$1 = (p+q)^n = \sum_{x=0}^{n} C_x^n p^x q^{n-x}$$

Binomial Mean

X is a binomial random variable with parameters p and n

Mean:

$\mu = E(X) = np$

$$\mu = \sum x C_x^n p^x q^{n-x} = p \frac{\partial}{\partial p} \sum C_x^n p^x q^{n-x} =$$

$$= p \frac{\partial}{\partial p} (p+q)^n = np$$

$$E(X(X-1)) =$$
$$= \sum x(x-1) C^n_x p^x q^{n-x}$$
$$= p^2 \frac{\partial^2}{\partial p^2} \sum C^n_x p^x q^{n-x} =$$
$$= p^2 \frac{\partial^2}{\partial p^2} (p+q)^n \Big|_{q=1-p} = n(n-1)p^2$$
$$E(X^2) = E(X(X-1)) + E(X) =$$
$$= n^2p^2 - np^2 + np = n^2p^2 + np(1-p)$$

$$V(X) = E(X^2) - E(X)^2 = n^2p^2 + np(1-p) - (np)^2$$
$$= \boxed{np(1-p)}$$

Binomial mean, variance and standard deviation

Let X be a binomial random variable with parameters p and n

- Mean:

$\mu = np$

- Variance:

$\sigma^2 = V(X) = np(1-p)$

- Standard deviation:

$\sigma = \sqrt{np(1-p)}$

- Standard deviation to mean ratio

$$\sigma/\mu = \sqrt{np(1-p)}/np = \frac{\sqrt{(1-p)/p}}{\sqrt{n}}$$