**The product rule**

P(A, B) = P(A)P(B|A) “ , ” means “and”

P(A, B) = P(B)P(A|B) “ | ” means “given”

P(A, B | C) = P(A | C)P(B | A, C)

**The sum rule**

P(A v B) = P(A) + P(B) – P(A, B) “ v “ means “or”

If A and B are mutually exclusive i.e. P(A, B) = 0

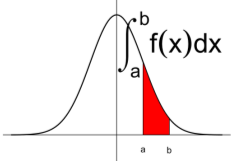
then P(A v B) = P(A) + P(B)

P(A v B | C) = P(A|C) + P(B|C) – P(A, B|C)

**Random Variables**

A random variable is a quantity that has an associated probability distribution

Discrete R.V.s Continuous R.V.s

Probability mass function Probability density function

**Log and exponents**

Logs and exponentials will come up from time to time. Be familiar with them. Note: e ~ 2.71828 and if I write log, I mean natural log

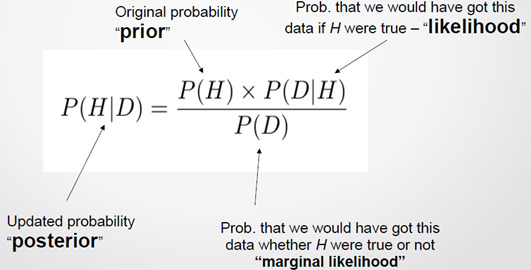
eaeb = ea + b

(ea)b = eab

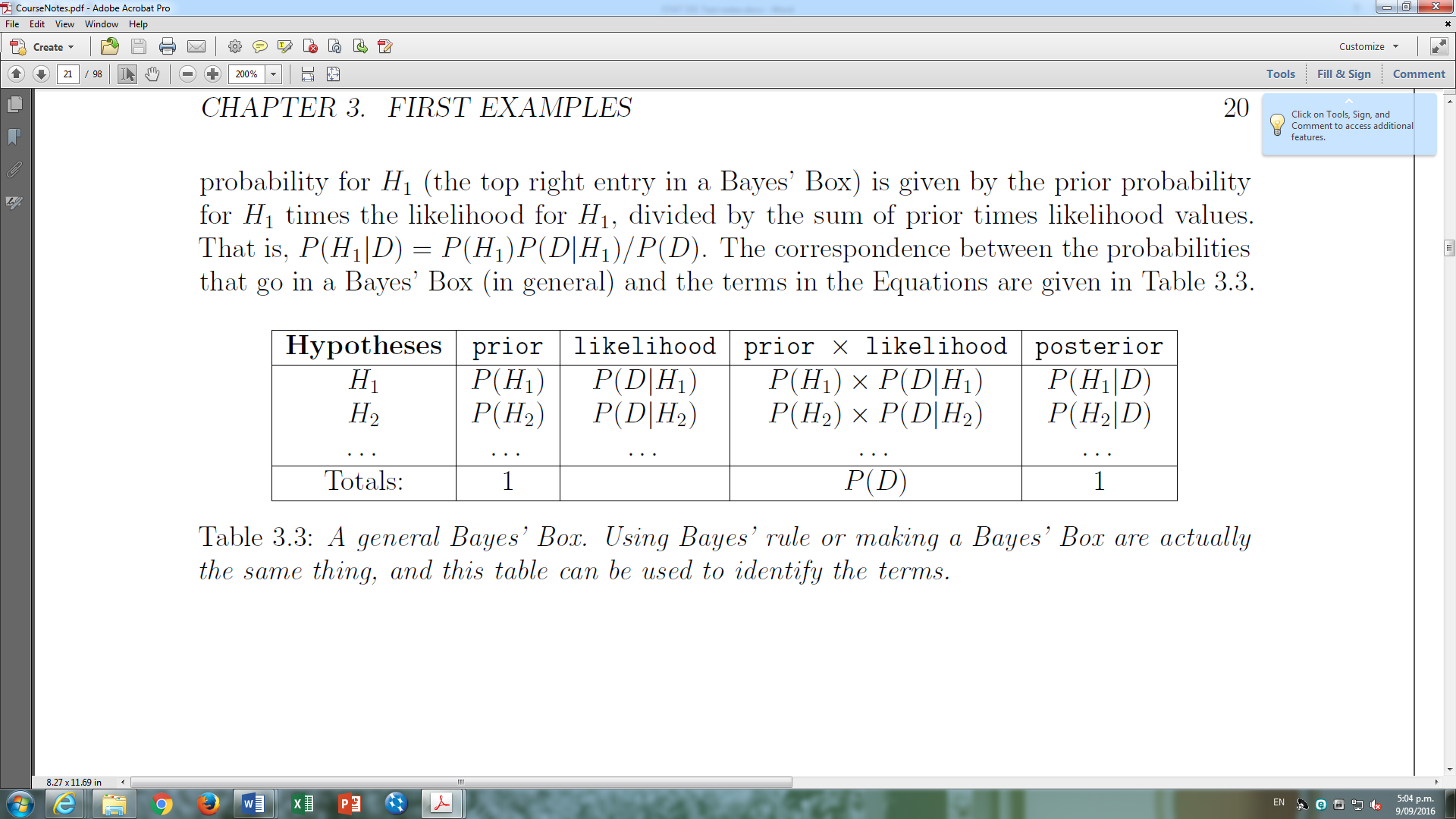
log(ab) = log(a) + log(b)

log(ab) = b log(a)

**Bayes' Rule**



**Bayes' Box**



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Theta | Prior | Likelihood | h | Posterior |
| θ | P(θ) | P(x|θ) | P(θ) P(x|θ) | P(θ|x) |
|  |  |  |  |  |
|  |  |  | P(x) |  |

**Marginal likelihood**

P(B) marginal likelihood is the probability of obtaining the data B without assuming that A is either true or false.

**Bayes' Rule for Sets of Hypotheses**

P(Hi|D) = P(Hi) P(D|Hi)

. P(D) .

P(D) = ΣNi = 1 P(Hi) P(D|Hi)

**Forms of Bayes rule**

p(θ|x) ᴕ p(θ)p(x|θ)

. p(x) .

p(θ|x) ᴕ p(θ)p(x|θ)

Posterior ᴕ prior × likelihood

**Discrete Parameter Estimation**

Bayes' rule can be applied to a set of hypotheses about the value of an unknown parameter

- A hypothesis might be “θ = 1”

- Another might be “θ = 2”

- Suppose we observed data “ x = 3”

Posterior distribution = Prior distribution \* Likelihood

Common normalisation constant

(or Marginal Likelihood)



**Prior distributions**

Prior for θ was uniform distribution between 0 and 1

θ ~ Uniform(0, 1)

If we knew the value of θ, x would have a binomial distribution with N trials and success probability θ.

x ~ Bin(N, θ)

**Posterior distributions**

Beta distribution is a value that is between 0 and 1, just like our θ. It has two parameters, α and β.

x ~ Beta(α, β)

Beta(hits,misses)

Beta(α, β) = ᶘ10 xα – 1 (1 – x) β – 1 dx

If the prior on an unknown parameter \_ is

x ~ Beta(α, β)

and the likelihood is

θ|x ~ Bin(θ, N)

then the posterior is given by

X|θ ~ Beta(α + X, β + N - X)

**Probability density function**

If y ~ Beta(α,β) for some positive α > 0 and β > 0, then the probability density function is given by

p(y) ᴕ yα – 1 (1 – y) β – 1 for y [0,1]

The expected value (or mean) and variance of y are given by

E(y) = α / (α+β)

var (y) = αβ / ((α+β)2(α+β+1))

Beta(3,17) = y2(1 – y) 16

Exp(Beta(3,17)) = 3 / 20

sdv(Beta(3,17)) = sqrt (3\*17 / ((3+17)2(3+17+1)))

= sqrt (51/ 8400) = 0.00607

**Point estimaste**

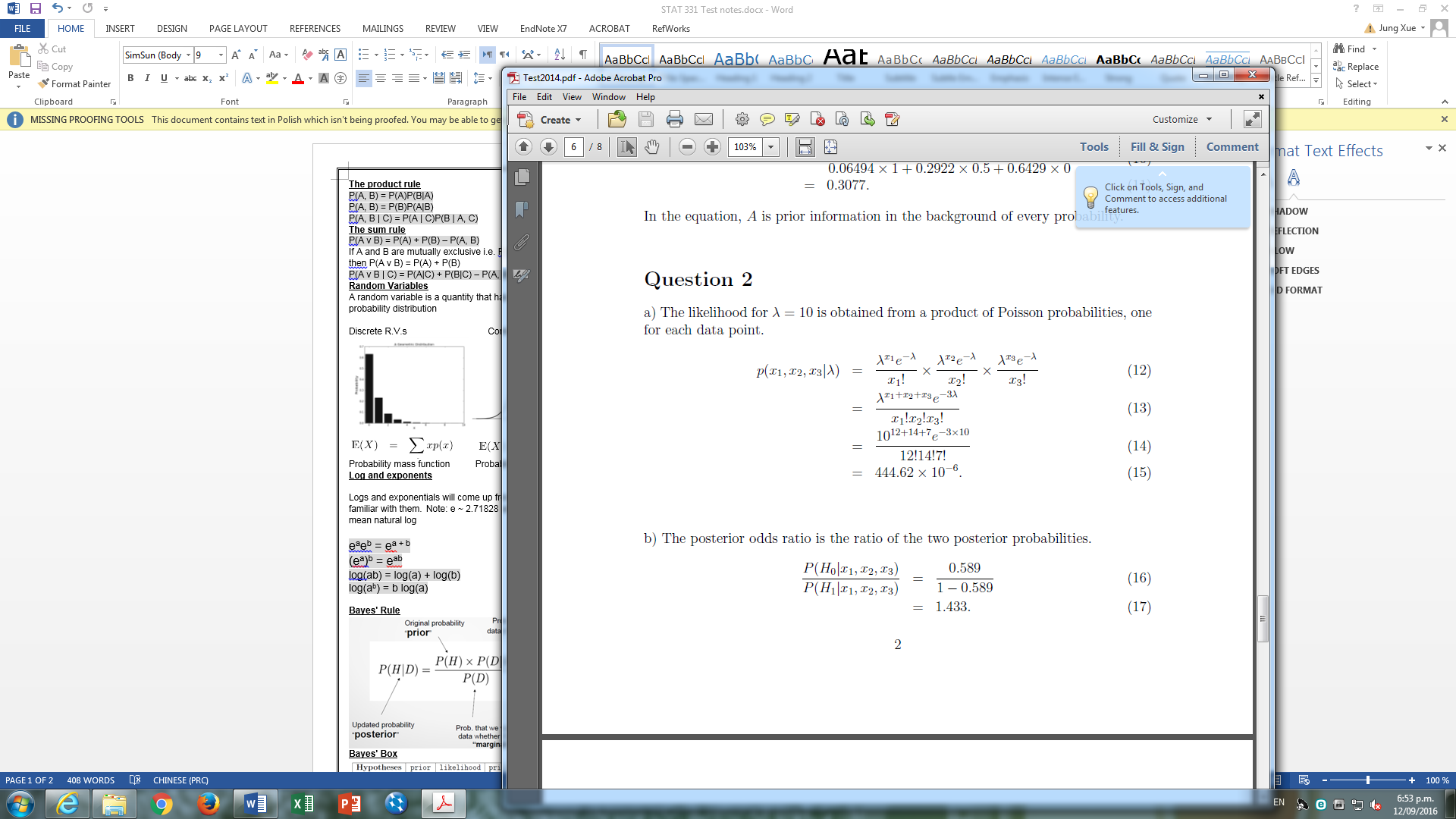
θ = pos mean ± pos std dev

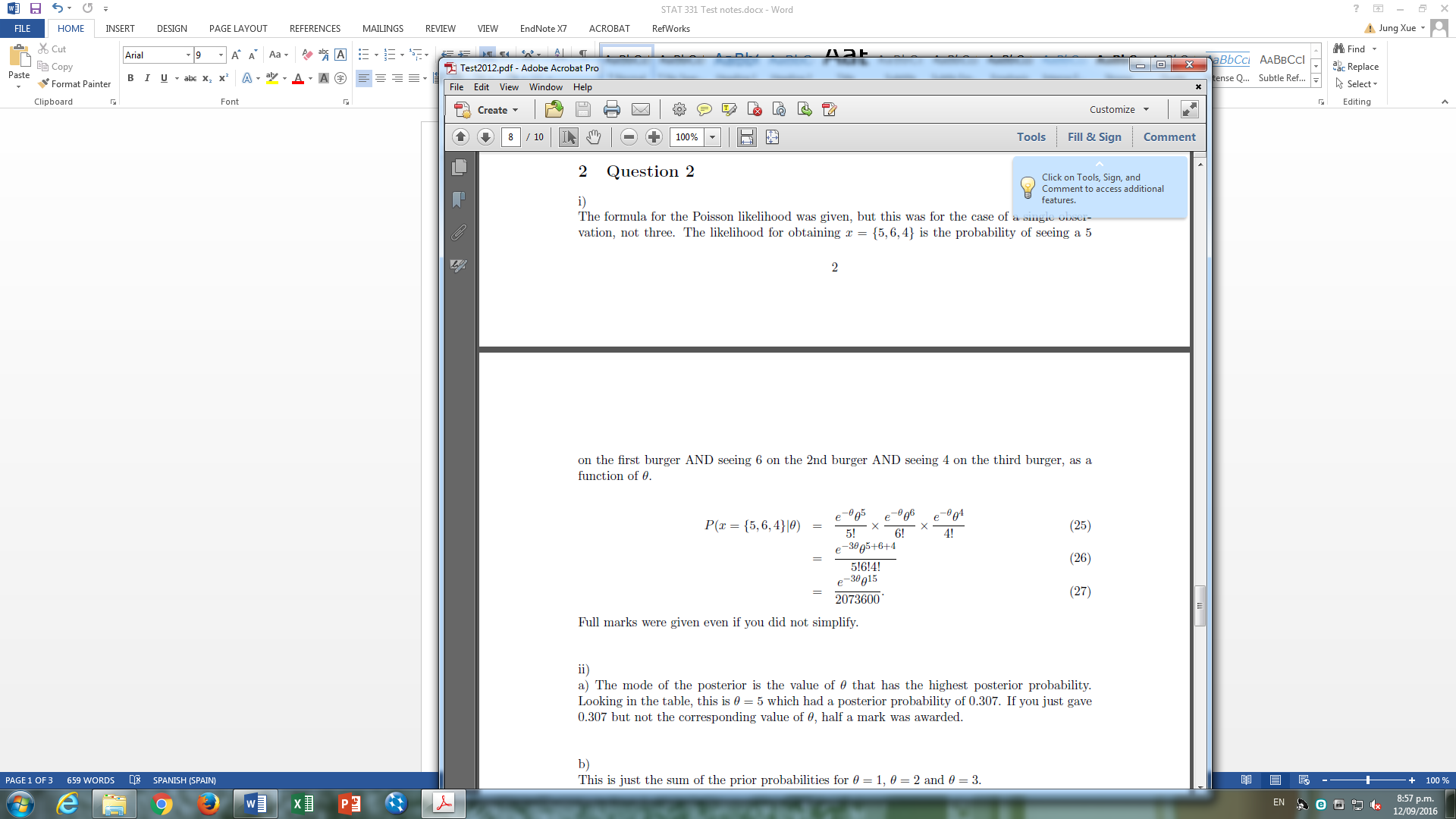
**Poisson distribution**



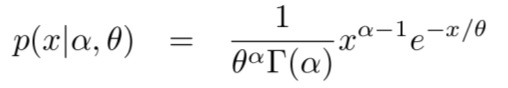
Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event.

independence of the observations.

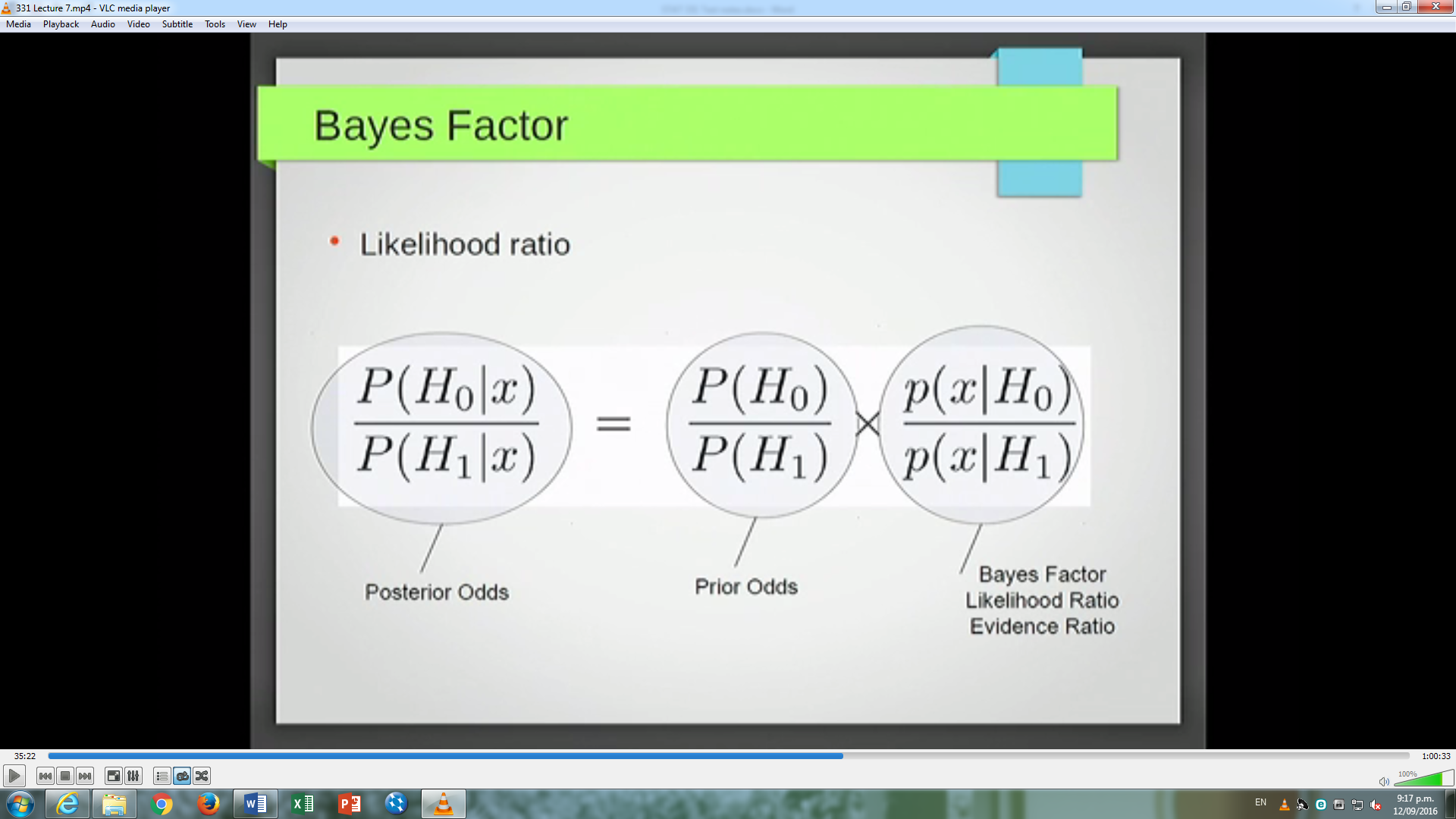




**Gamma**



**Hypothesis testing**



**Metropolis algorithm**

[2 marks] If the Metropolis algorithm was used to sample the posterior distribution under H1 in this problem, what would be the acceptance probability for a proposed move from = 145 to = 146?

post[2]/post[1]=1.348485, thus 1.

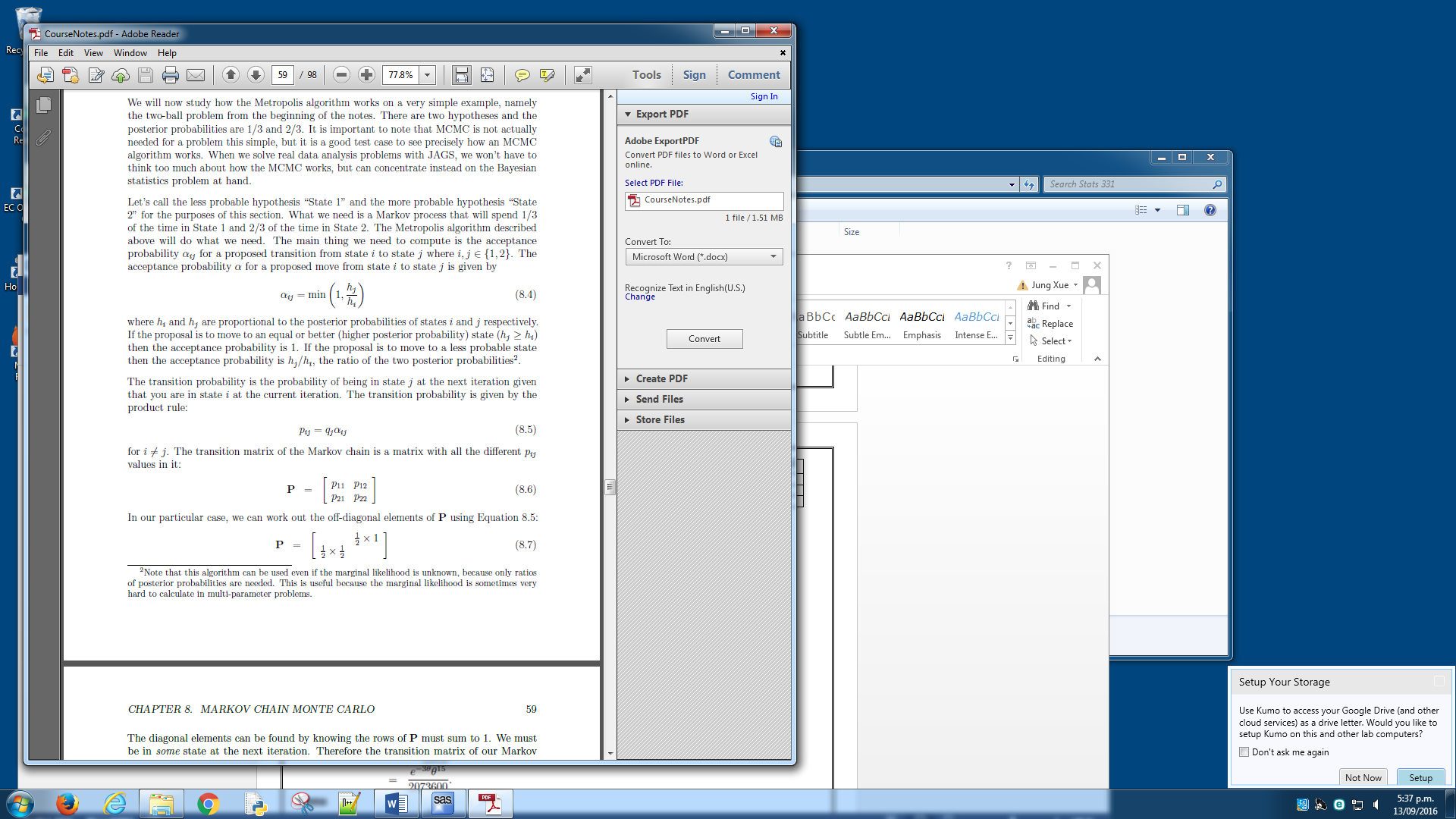
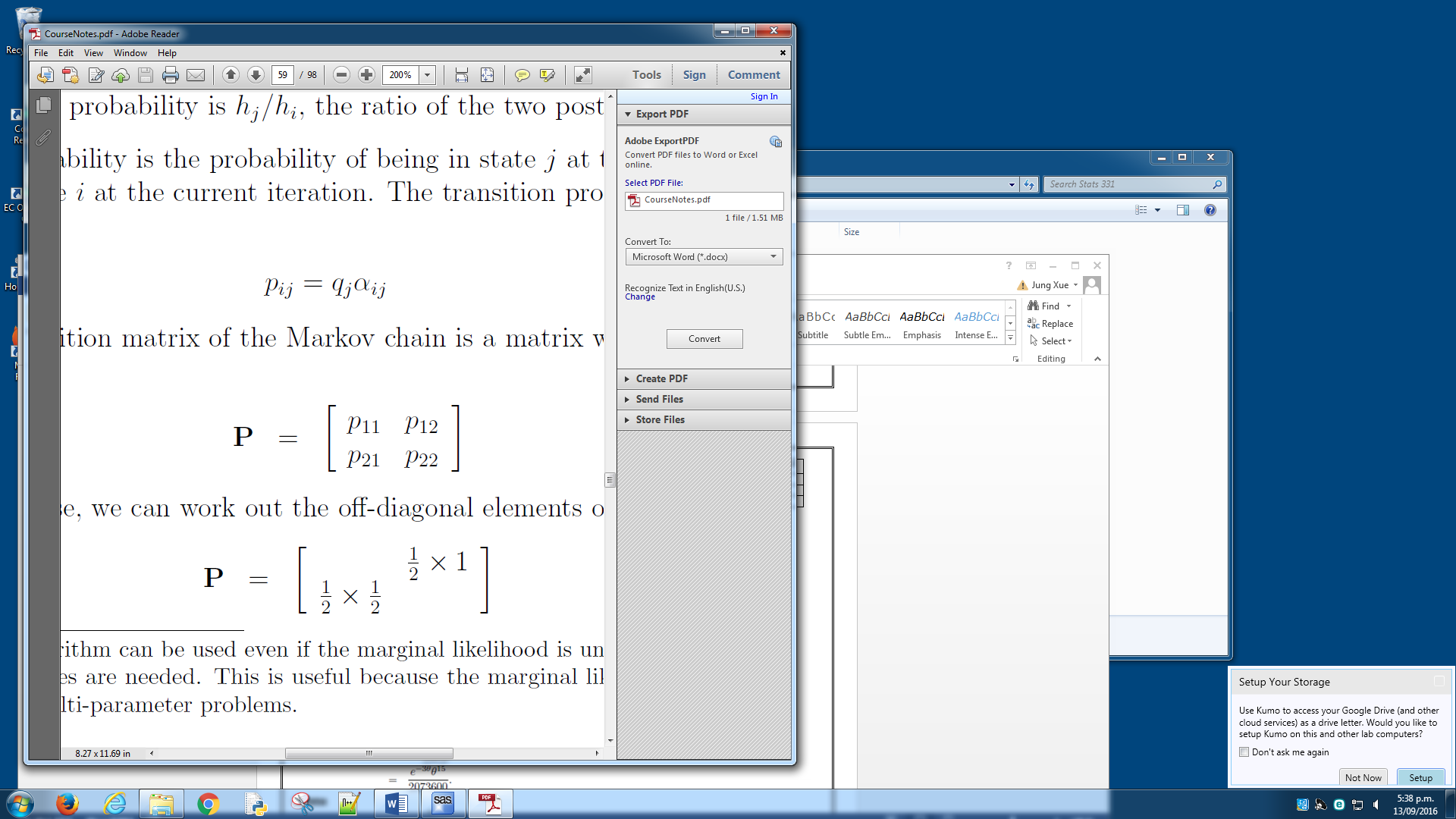
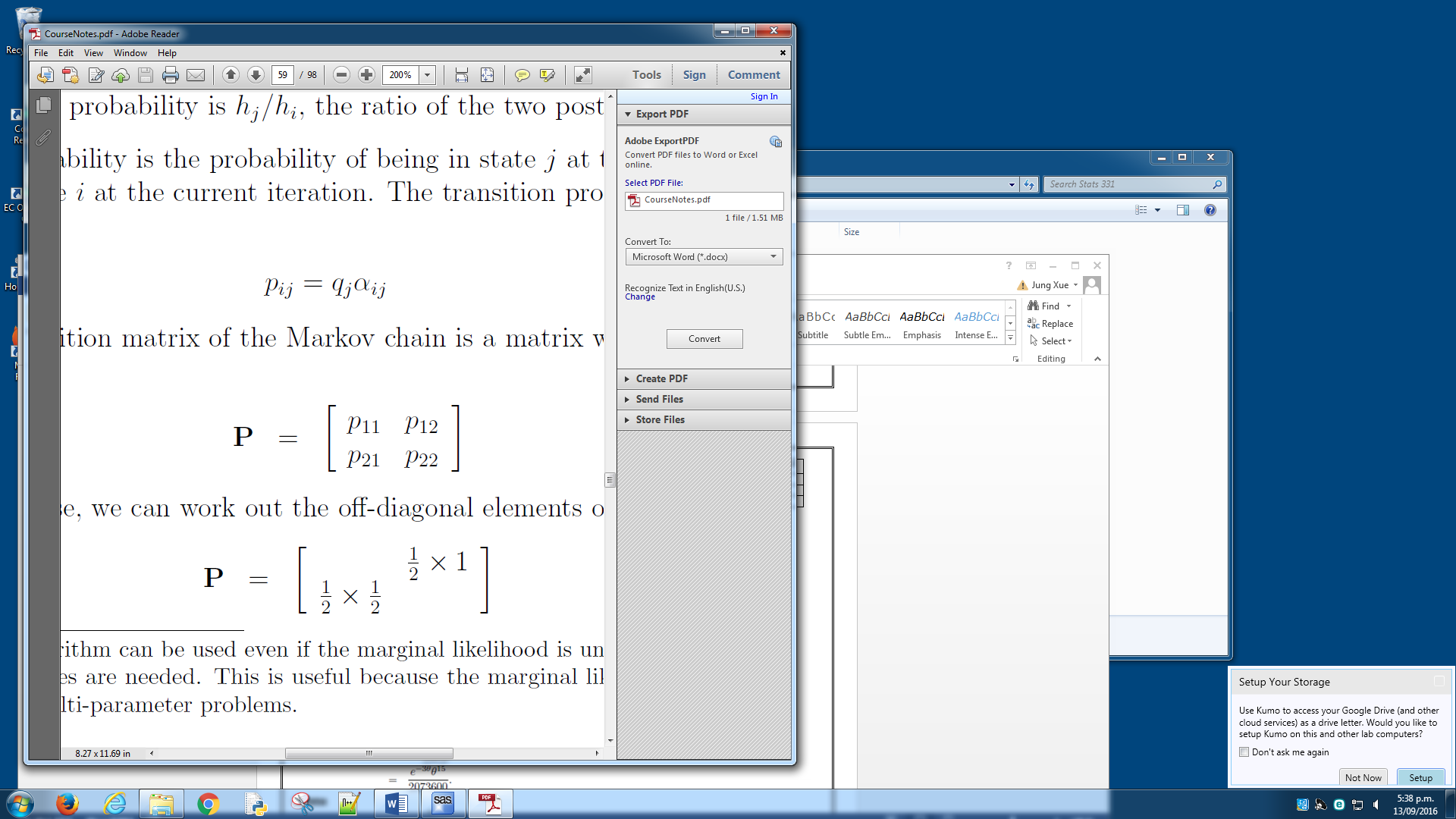
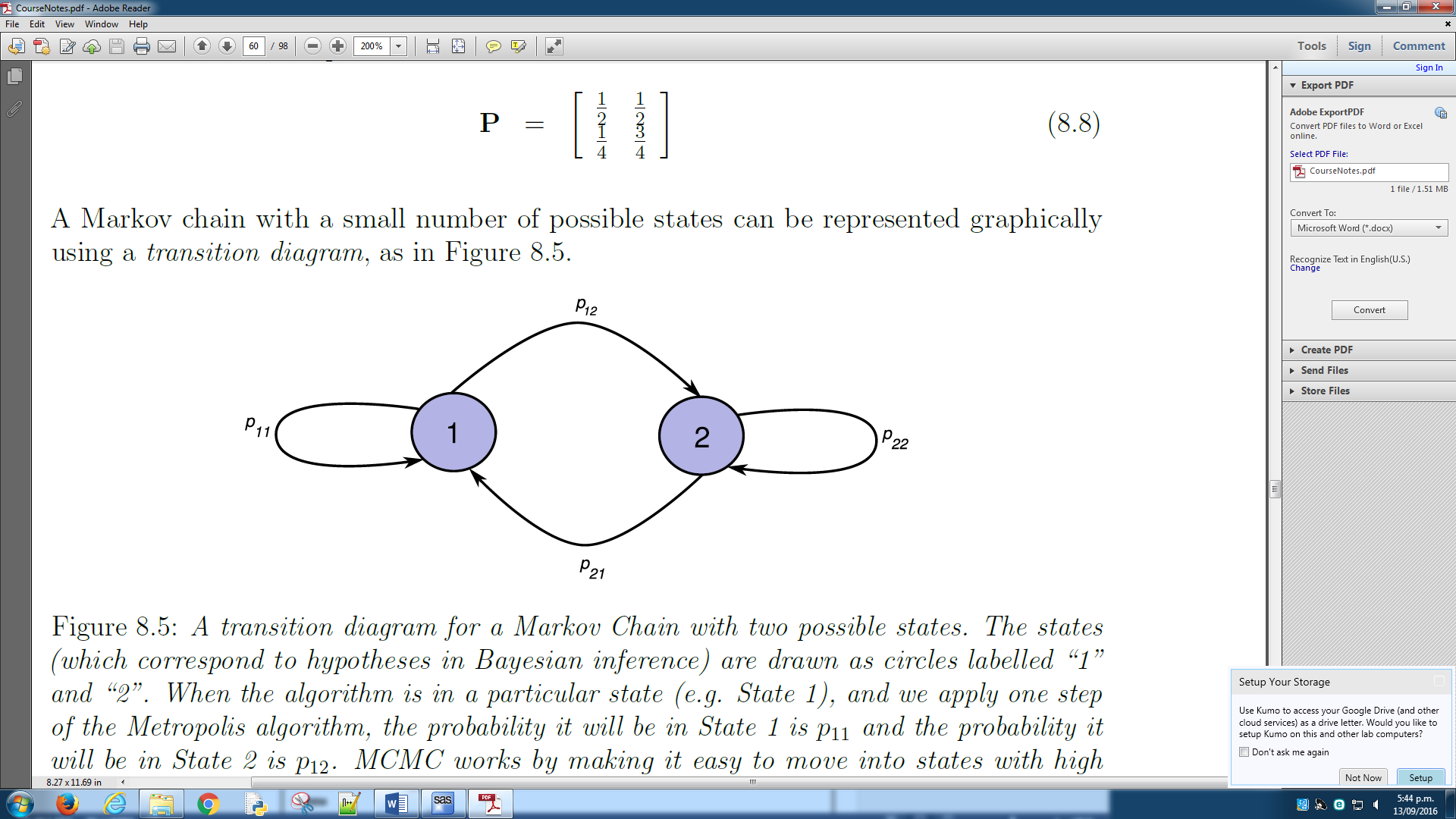
f) [2 marks] If the Metropolis algorithm was used to sample the posterior distribution under H1 in this problem, what would be the acceptance probability for a proposed move from = 160 to = 170?

post[26]/post[16]=0.4064959

Acceptance probability

The acceptance probability is 1 if hj ≥ hi (i.e. if moving to a more probable state) and hj/hi otherwise (if moving to a less probable state).

**MCMC Two State Problem**

**Symbol**

π = multiple product of 1 - i

Σ = sum of 1 - i