BLCMs for continuous tests

CA18208 HARMONY Serbia Training School - https://harmony-net.eu/

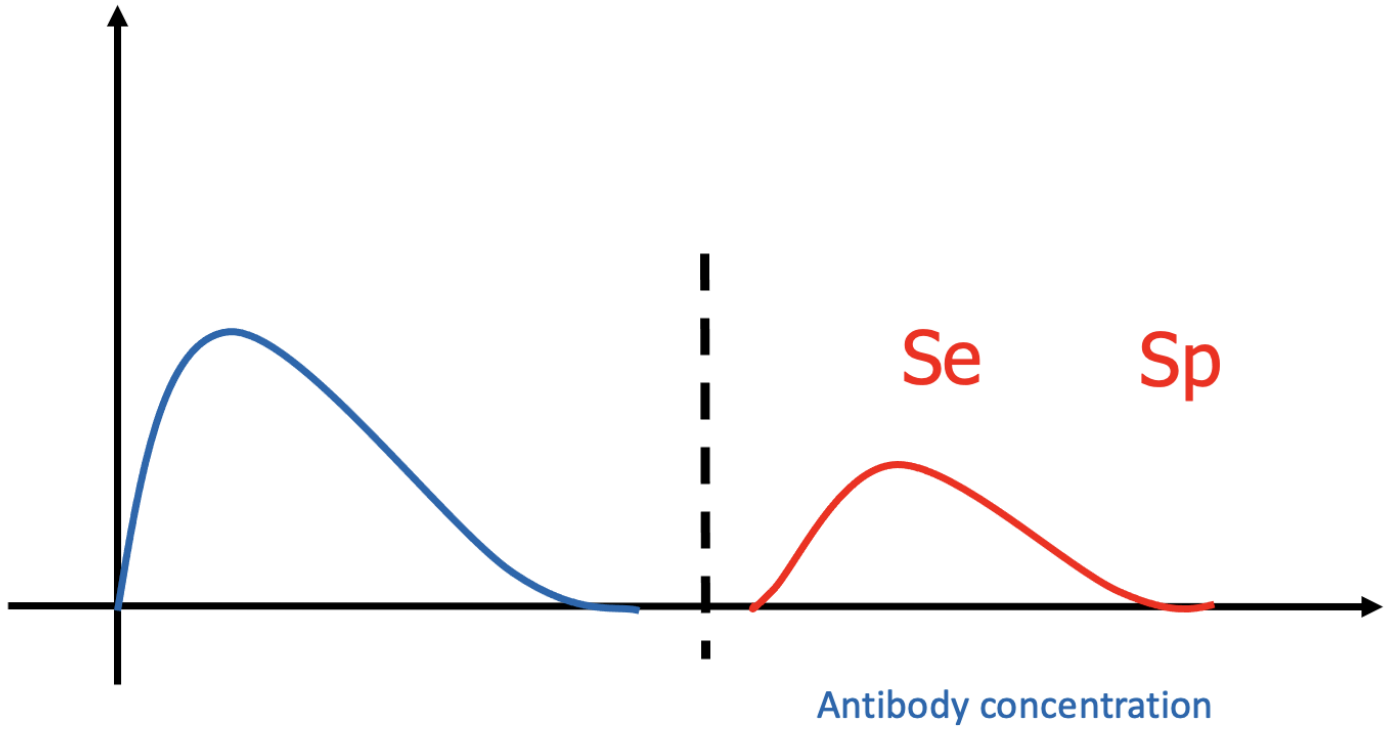
Eleftherios Meletis Julio Alvarez

2022-09-01

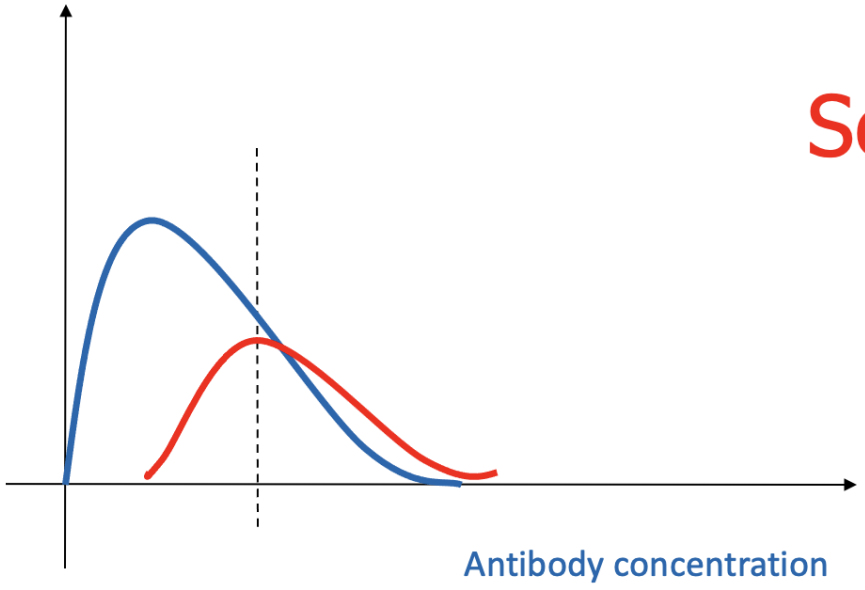
Presentation Outline

- Didactic Teaching
  - Receiver Operating Characteristic curve (ROC) analysis
  - ROC Analysis with BLCMs
- Practical Session
  - Hands-on example

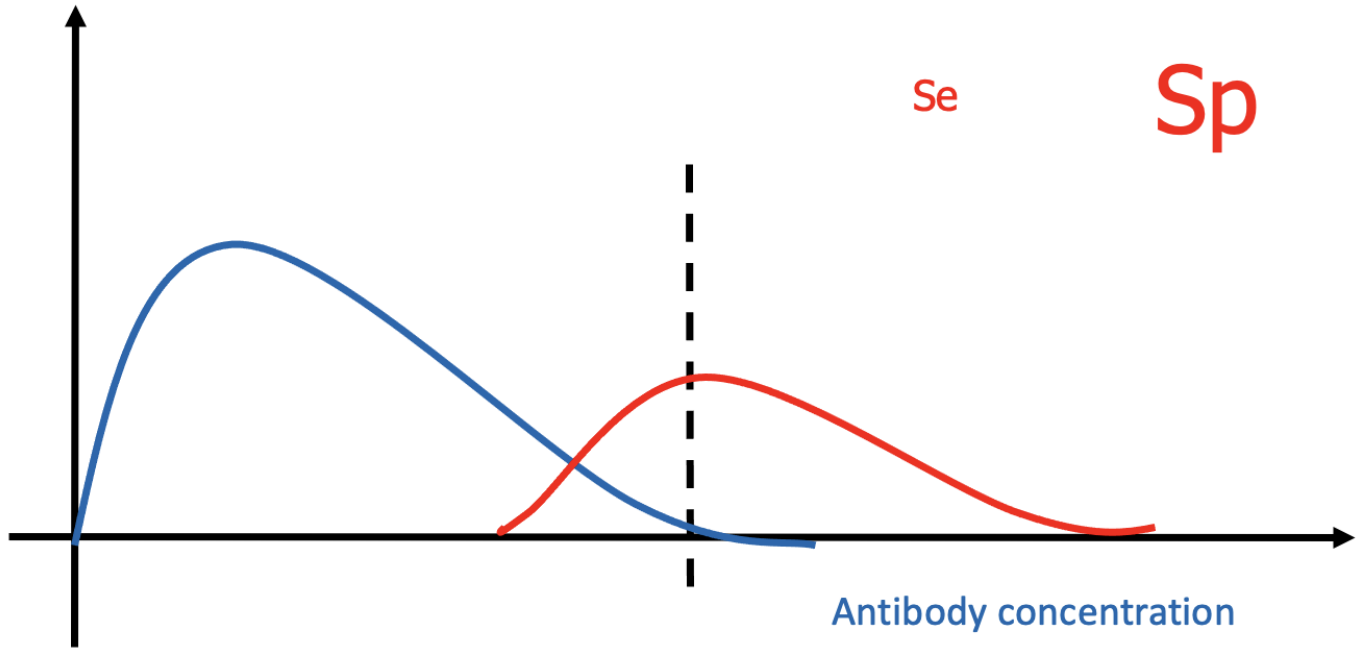
What type of test (Se,Sp) does this picture describe?



On the other hand this picture describes a test with poor discriminatory ability, right?

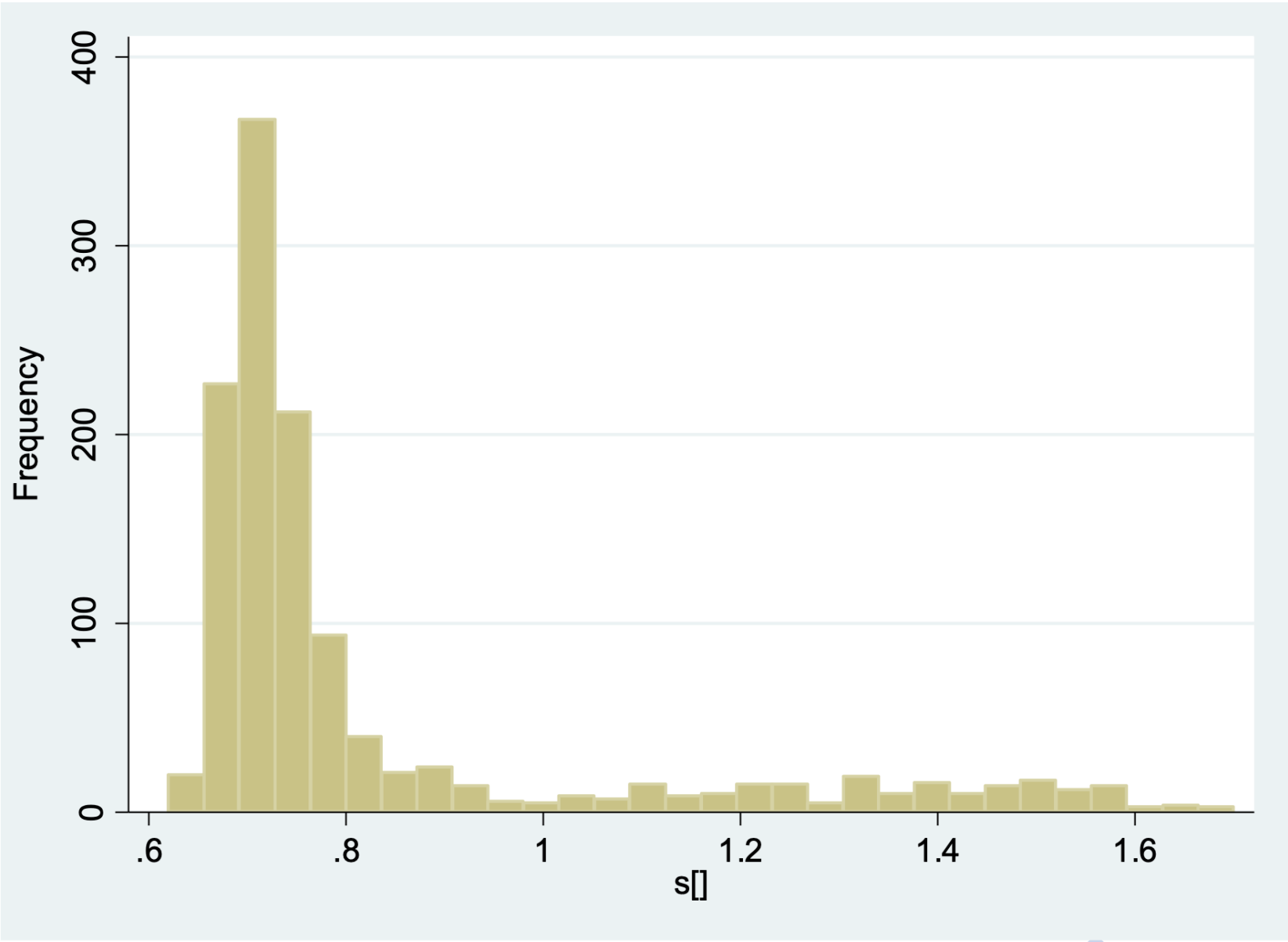


The following picture describes the most usual setting.



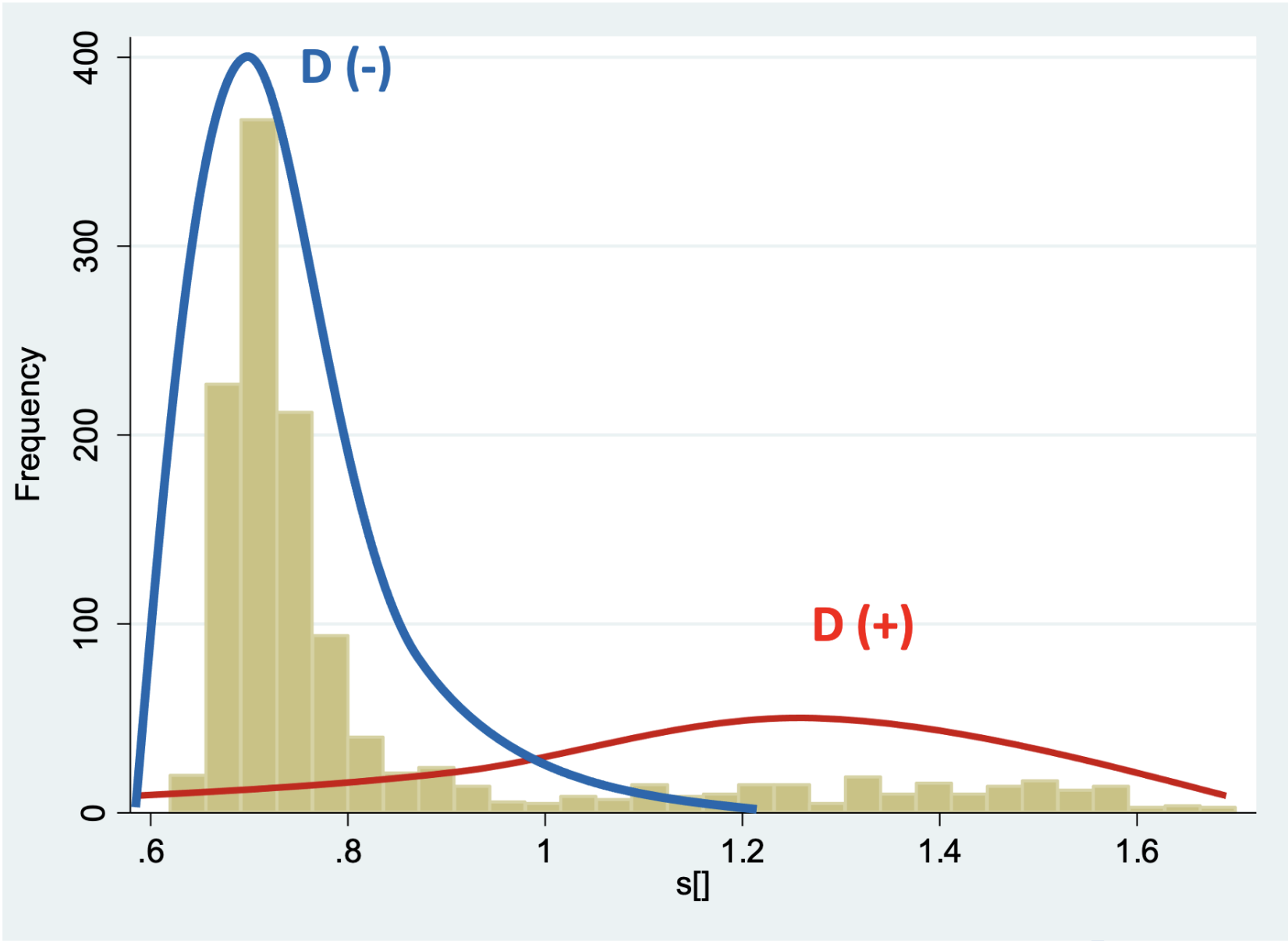
Let's start building our hands-on example.

In the next slide we'll see a histogram of the values of a continuous test result (e.g. ELISA measuring antibodies), on the logarithmic scale.



Antibody concentration

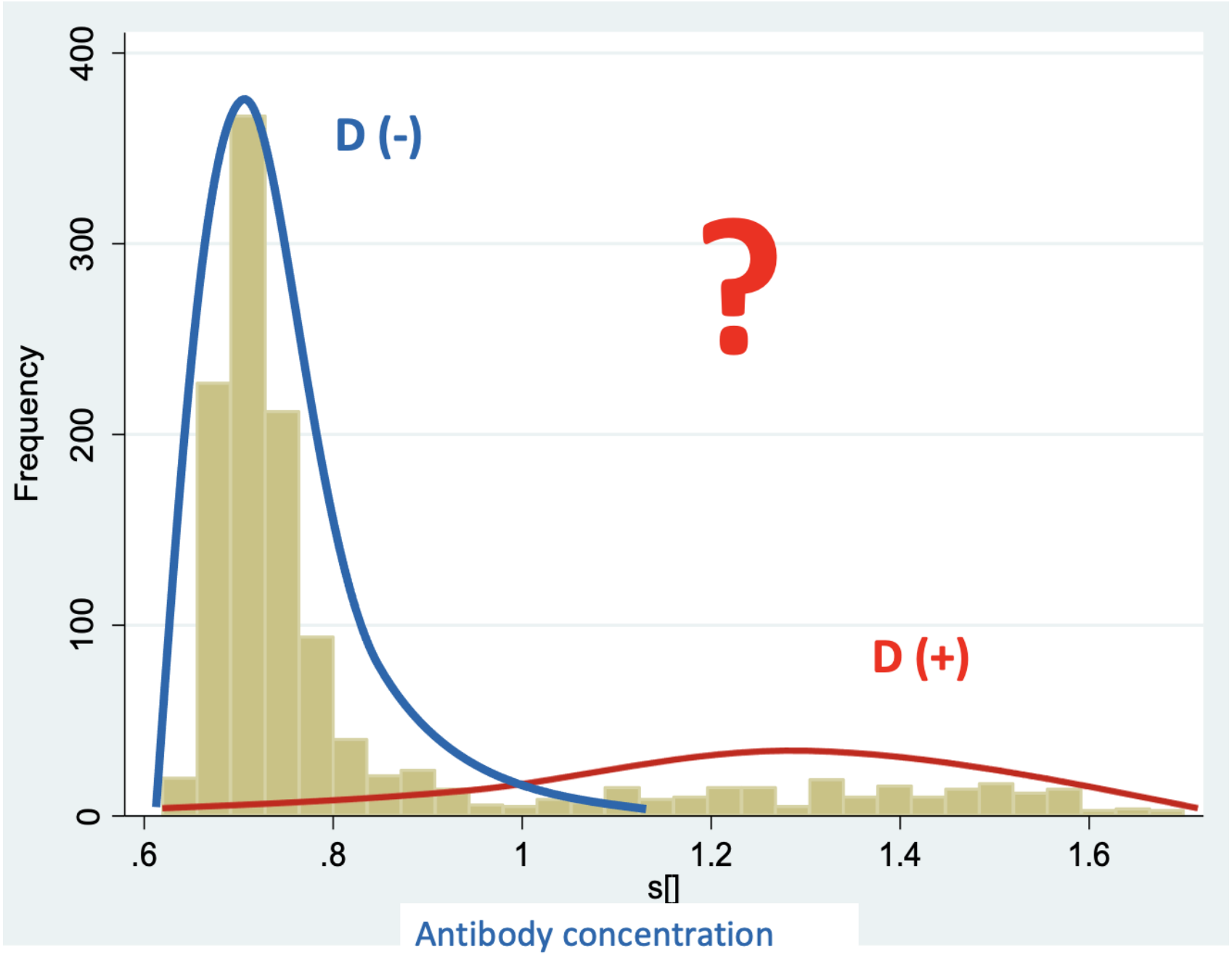
Assuming the infectious status of each individual known we can plot/add the distributions of the diseased and healthy.



ROC Analysis with BLCMs

Point 1

- In the absence of a gold standard the true disease status of each individual is unknown.
- The available information we have is the continuous test output from each individual.


D (-)
?
D (+)
Frequency
0
100
200
300
400
.6
.8
1
1.2
1.4
1.6
s[]
Antibody concentration

Point 2

- Most studies dichotomize the continuous test output based on a pre-selected cut-off value and apply the BLCMs we discussed yesterday.
- But this results in loss of valuable information.
  - All positive results are equal, no matter how near or far they are from the cutoff.

Model Specification - Mixture Normal Model

- The data are best described by a mixture of two normal distributions:
  - D (-) individuals with mean (mu1) and variance (1/tau_1)
  - D (+) individuals with mean (mu2) and variance (1/tau_2)
    - 1/tau = Precision
- The disease status for each indivudual is indicated by a latent variable.
- For identifiability we assume: $mu1 < mu2$
  - Diseased individuals are expected to have higher value of the continuous marker

Mixture Normal Model explained

```
model {
  for (i in 1:481) {
        #S[i] diagnostic test value for ith individual
        S[i] ~ dnorm(mu[i],tau[i])

        #Value of mu & tau depending on the group (diseased or disease-free)
        mu[i] <- lambda[T[i]]
        tau[i] <- gamma[T[i]]
        #dcat <- categorical #D(-) if T[i]=1, D(+) if T[i]=2
        T[i] ~ dcat(P[])
  }
    P[1:2] ~ ddirch(alpha[])

    # lambda[1]-gamma[1] mean-precision of non-disease group
    lambda[1] ~ dnorm(0,0.001)
    lambda[2] ~ dnorm(0,0.001)T(lambda[1],)
    gamma[1] ~ dgamma(0.001,0.001)
    gamma[2] ~ dgamma(0.001,0.001)

    # variance = 1/precision(tau)
    sigma[1] <- 1/gamma[1]
    sigma[2] <- 1/gamma[2]

    # AUC
    AUC <- phi(-(lambda[1]-lambda[2])/sqrt(sigma[2]+sigma[1]))
    # ROC curve
```

Data - Initial Values

```
summary(S)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -3.100  -2.900  -2.700  -2.414  -2.430   0.840
```

Define initial values:

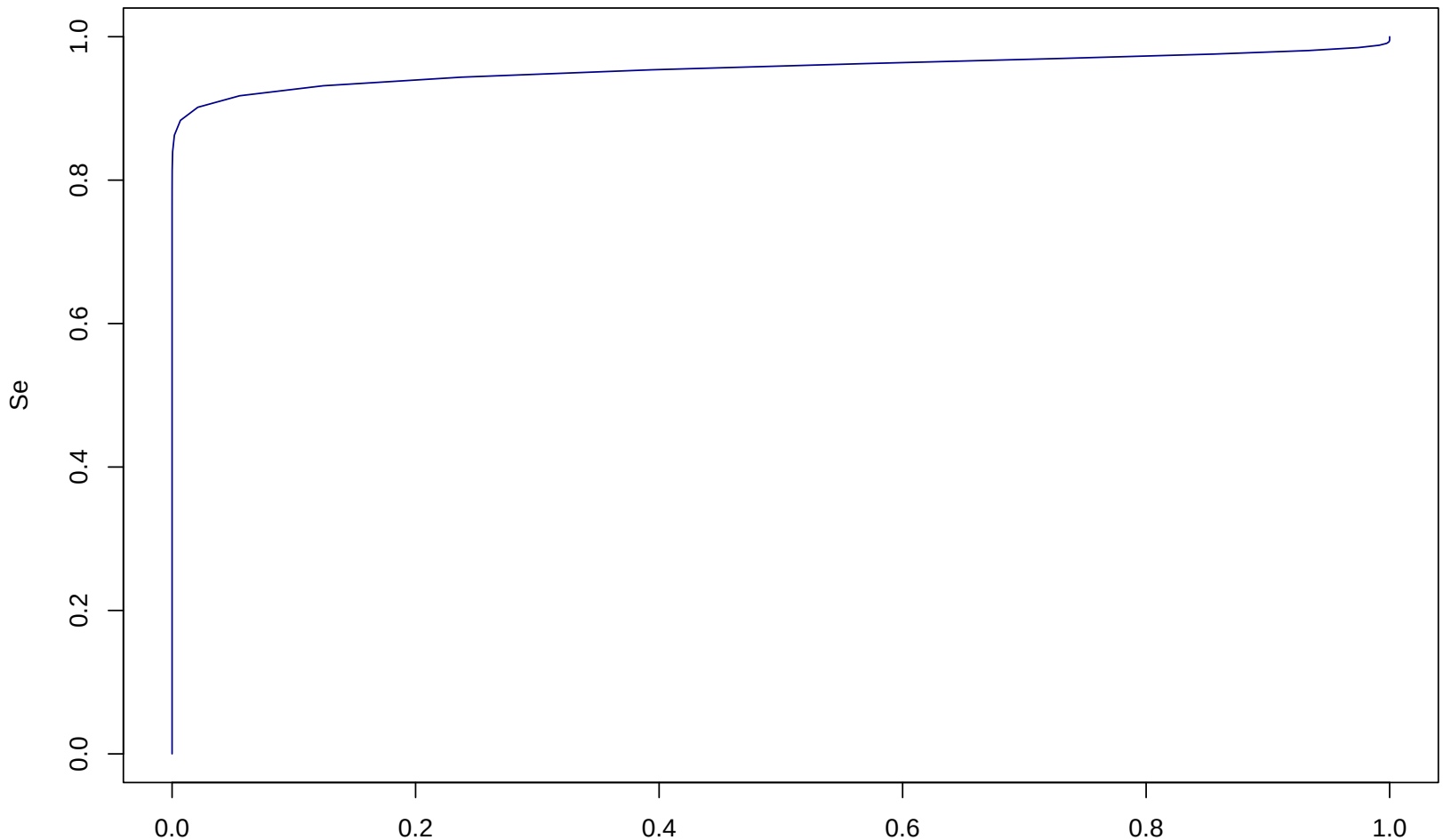
```
lambda <- list(chain1=c(-3, 0), chain2=c(-2,-2))
gamma <- list(chain1=c(10, 0.1), chain2=c(30, 5))
```

```
results <- run.jags('cont_test.txt', n.chains = 2)
```

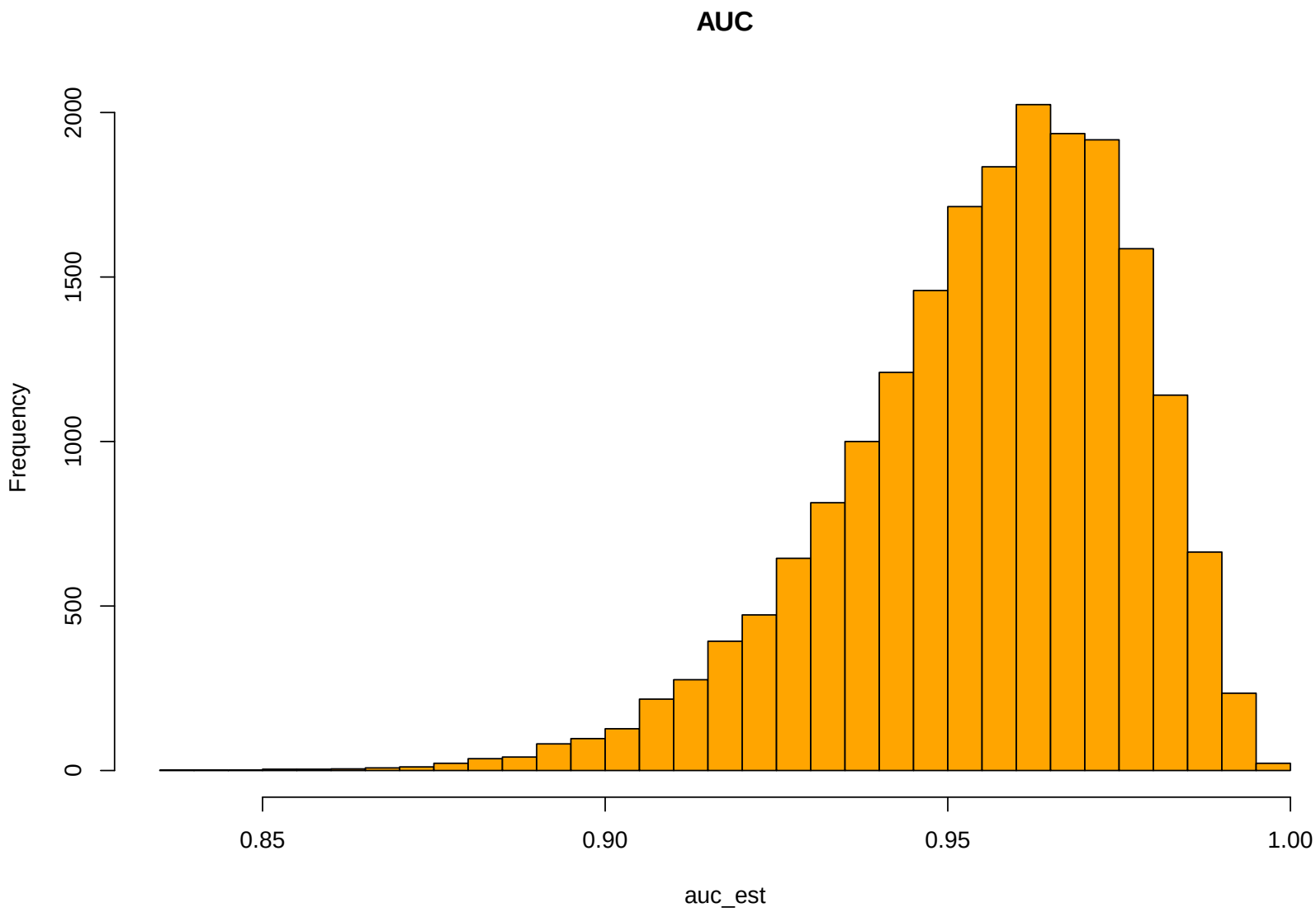
```
#plot(results, vars=c('AUC', 'P', 'lambda', 'gamma', 'sigma'))
```

```
results_summary <- add.summary(results, vars=c('AUC', 'P', 'lambda', 'gamma', '
```

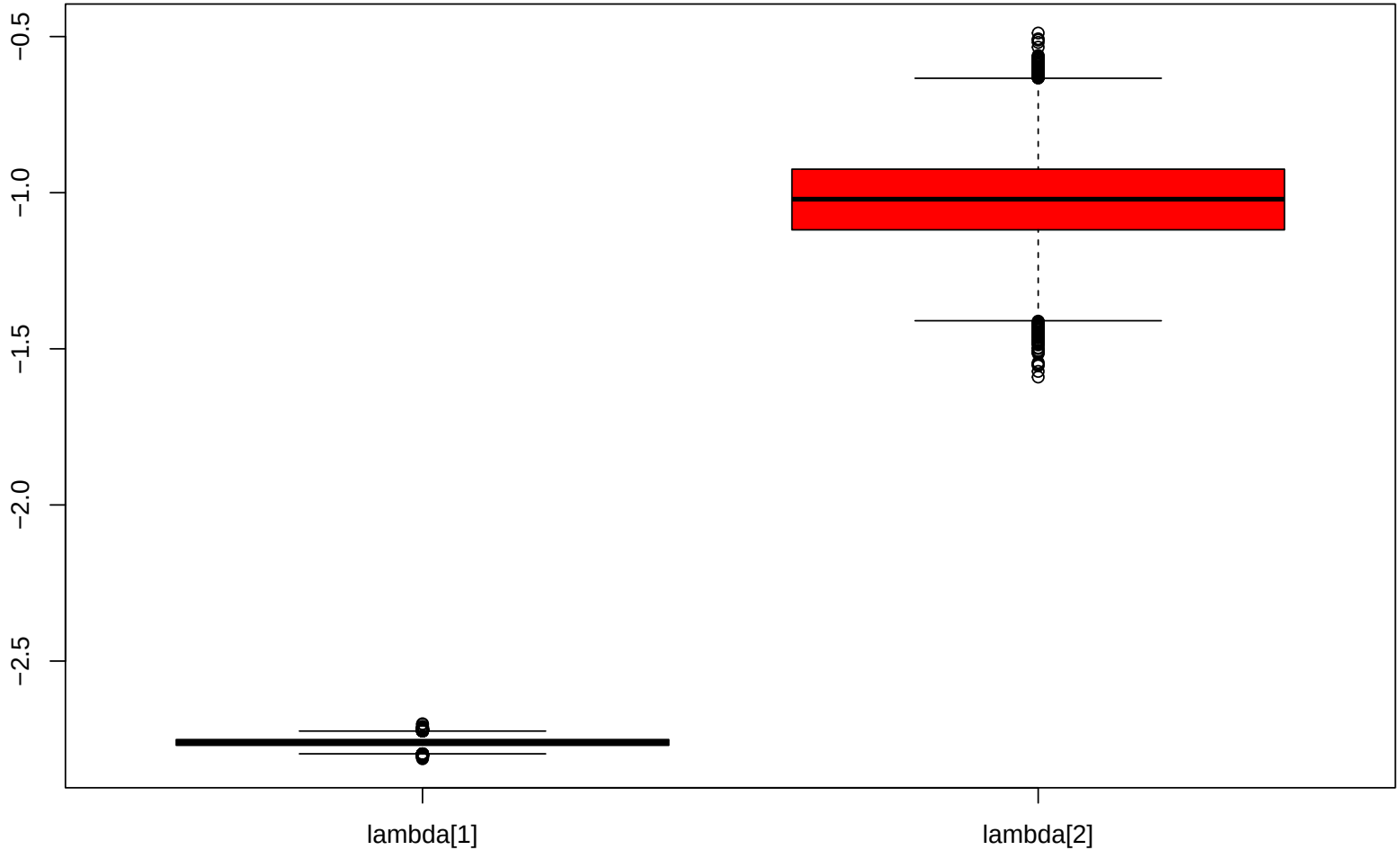
```r
se_est <- combine.mcmc(results, vars='se')
sp_est <- combine.mcmc(results, vars='sp')
ses_mu <- apply(se_est, 2, mean)
sps_mu <- apply(sp_est, 2, mean)
par(mfrow=c(1,1))
plot((1-sps_mu), ses_mu, type="l", col="darkblue", xlab = "1-Sp", ylab = "Se")
```



```
auc_est <- combine.mcmc(results, vars='AUC')
hist(auc_est, breaks=50, col="orange", main="AUC")
```



```
lambda_est <- combine.mcmc(results, vars='lambda')
boxplot(as.matrix(lambda_est), col="red")
```



Conclusion - Remarks

- Normality assumption?
- Distance between the D(-) and D(+) distributions
- Label switching
- More complicated settings
  - Correlated tests
  - Multiple populations
  - More than 2 infectious stages
  - etc...

Exercises

- Run the model and produce the same output.
- Try to run the model under different prior specification for lambda
- What happens if you remove T(lambda[1],)? Does the model converge?
- Try and find the cut-off value that maximizes Youden's index?

Another approach

```
# Another option in JAGS is to use dnormmix:
# S[i] ~ dnormmix(mu[1:2], tau[1:2], P[1:2])
# This is more efficient than explicitly simulating the latent class
#modules# mix
#factories# mix::TemperedMix sampler off
##### /Alternative
```