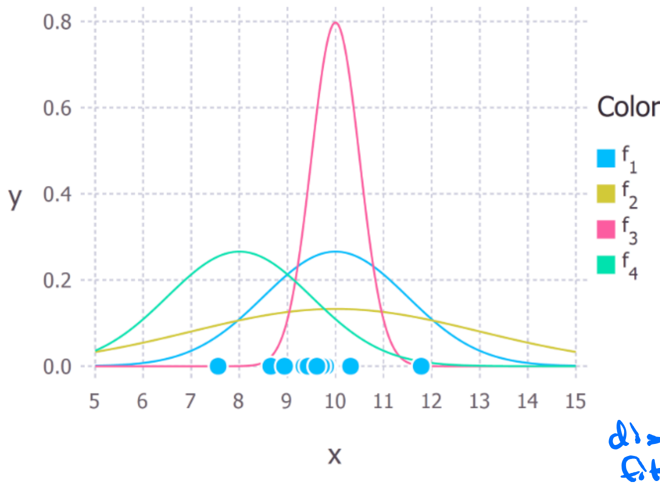
Stats 135, Fall 2019

Lecture 9, Wednesday, 9/19/2019

CLASS ANNOUNCEMENTS: RStudio demonstration in-class today.

stat 135 lec 9



Which normal distribution corresponds to our data?

answer

$f_1 = N(10, 2.25)$

$\hat{\mu}_{ML} = 10$

$\widehat{\sigma^2}_{ML} = 2.25$

The ML estimator is the distribution parameters that best fit your data.

Last time For $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

ML estimator

$$\hat{\mu} = \bar{X}$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n} S^2$$

$$\frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2_{n-1}$$

$$\Rightarrow \frac{n\widehat{\sigma^2}}{\sigma^2} = \frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1}$$

Today

⓪ Sec 6.3 t-distribution

① Sec 6.5.3 Find 95% CI for μ, σ^2 for $N(\mu, \sigma^2)$

② Sec 8.5.3 Parametric bootstrap in R (to find 95% CI)

③ Sec 8.5.2 large sample properties of $\hat{\theta}_{ML}$ (to find 95% CI)

t-distribution (see chap 6)

$\underline{\text{def}^n}$ $Z \sim N(0,1)$

$U \sim \chi^2_{n-1}$

$$\frac{Z}{\sqrt{U/_{n-1}}} \sim t_{n-1}$$

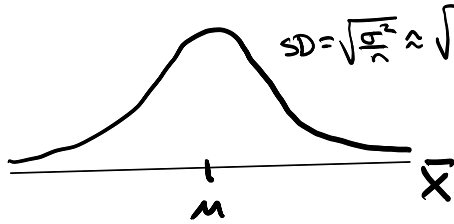
$\nwarrow$ d.f.

Let $x_1, \ldots, x_n \overset{iid}{\sim} N(\mu, \sigma^2)$ where μ, σ^2 are unknown.

The t-distribution comes up when making a CI for μ.

$\bar{X}$ has sampling distribution

$SD = \sqrt{\frac{\sigma^2}{n}} \approx \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$



"normalize $\bar{X}$"

Note $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

but $\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

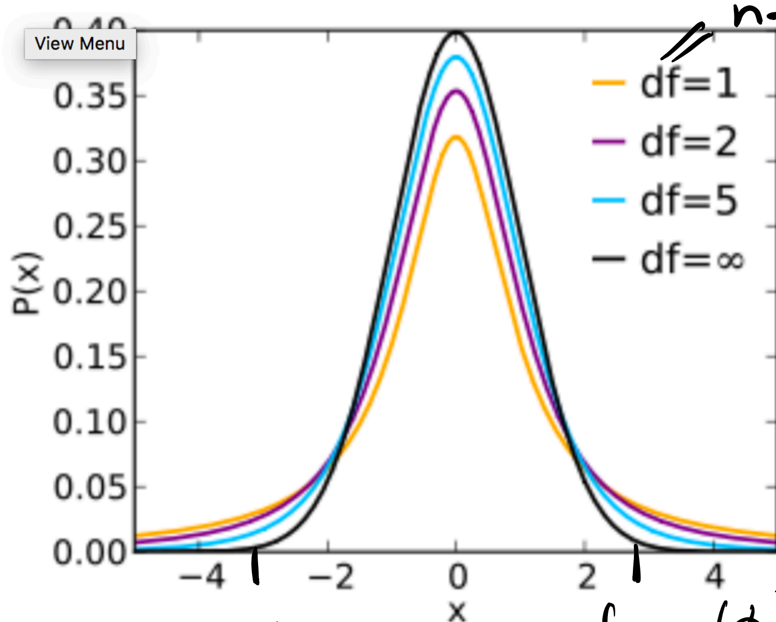
$$\frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)s^2}{\sigma^2} \Big/ n-1}} \sim t_{n-1}$$

$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = Z$

$\leftarrow \sqrt{U/_{n-1}}$

$U \sim \chi^2_{n-1}$

$n-1$



$t_{n-1}\left(\frac{\alpha}{2}\right)$ ← $\frac{\alpha}{2}$ quantile of t_{n-1} $\qquad t_{n-1}\left(\frac{\alpha}{2}\right)$

Sec 8.5.3 95% CI for μ and σ^2 of $N(\mu,\sigma^2)$ where $X_1,\dots,X_n \overset{iid}{\sim} N(\mu,\sigma^2)$

95% CI of μ:

we have,

$$P\left(\frac{\bar{X}-\mu}{S/\sqrt{n}} \in \left[-t_{n-1}\left(\tfrac{\alpha}{2}\right), t_{n-1}\left(\tfrac{\alpha}{2}\right)\right]\right) = 1-\alpha$$

$$\Rightarrow P\left(\mu \in \left[\bar{X} - t_{n-1}\left(\tfrac{\alpha}{2}\right)\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1}\left(\tfrac{\alpha}{2}\right)\frac{S}{\sqrt{n}}\right]\right) = 1-\alpha$$

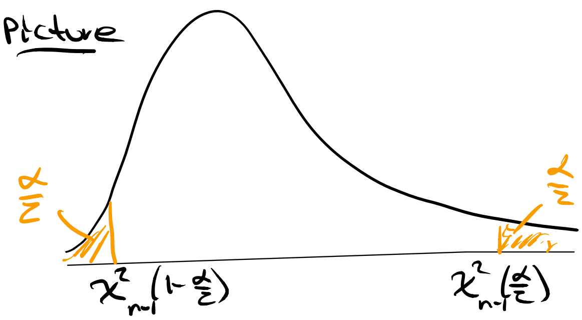
$(1-\alpha)100\%$ CI for μ.

Here $\frac{S}{\sqrt{n}}$ is an estimate of the SE of $\hat{\mu}_{ML}$.

Next 95% CI of σ^2:

Recall $\frac{n\widehat{\sigma^2}}{\sigma^2} = \frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1}$

Picture



$\frac{n}{\sigma^2}\widehat{\sigma^2} \sim \chi^2_{n-1}$

we have,

$$P\left(\frac{n}{\sigma^2}\widehat{\sigma^2} \in \left[\chi^2_{n-1}(1-\tfrac{\alpha}{2}), \chi^2_{n-1}(\tfrac{\alpha}{2})\right]\right) = 1-\alpha$$

$$\Rightarrow P\left(\frac{1}{\sigma^2} \in \left[\frac{1}{n\widehat{\sigma^2}}\chi^2_{n-1}(1-\alpha), \frac{1}{n\widehat{\sigma^2}}\chi^2_{n-1}(\tfrac{\alpha}{2})\right]\right) = 1-\alpha$$

Note that $3 \in [2,4] \Rightarrow \frac{1}{3} \in [\frac{1}{4}, \frac{1}{2}]$

so

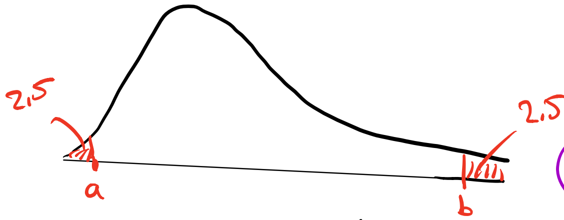
$$P\left(\sigma^2 \in \left[n\widehat{\sigma^2} \cdot \frac{1}{\chi^2_{n-1}(\frac{\alpha}{2})}, n\widehat{\sigma^2}\frac{1}{\chi^2(1-\frac{\alpha}{2})}\right]\right) = 1-\alpha$$

← $(1-\alpha)100\%$ CI for σ^2

② Sec 8.5.3 **Parametric bootstrap in R**

(to find 95% CI)

To make a 95% CI recall from lecture 5:



$\theta^* - \hat{\theta}$

$\hat{\theta} - \theta$

1000 MLE estimates θ^*

original MOM estimate. $\hat{\theta}$

let a, b be 2.5 and 97.5 percentile of $\hat{\theta} - \theta$

i.e $P(a < \hat{\theta} - \theta < b) = 95\%$

$\Leftrightarrow P(\hat{\theta} - b < \theta < \hat{\theta} - a) = 95\%$

This is a 95% CI of θ.

In R,

$b = \text{quantile}(\theta^* - \hat{\theta}, 1 - \frac{\alpha}{2})$

$a = \text{quantile}(\theta^* - \hat{\theta}, \frac{\alpha}{2})$

$[\hat{\theta} - b, \hat{\theta} - a] = \hat{\theta} - \underline{\text{quantile}(\theta^* - \hat{\theta}, c(1 - \frac{\alpha}{2}, \frac{\alpha}{2}))}$

$\hat{\theta}$ is a constant. $\rightarrow$ $\text{quantile}(\theta^*, c(1 - \frac{\alpha}{2}, \frac{\alpha}{2})) - \hat{\theta}$

$= 2\hat{\theta} - \text{quantile}(\theta^*, c(1 - \frac{\alpha}{2}, \frac{\alpha}{2}))$

1a) The parametric bootstrap of estimator SE

In a previous lab you saw the nonparametric bootstrap where you resample from the sample to approximate the se of some location parameter such as the sample mean. The parametric bootstrap may be used to find the se of a MOM or MLE estimate of a distribution parameter. Below we compare and contrast these approaches for finding the se of the MOM approximation to the Poisson rate parameter λ.

In the parametric bootstrap we find a sample then we look at the distribution of the sample and guess a model. We use MOM (or MLE) to find estimator of parameters. We repeatedly simulate model with estimated parameters and find se of sampling distribution.

Step 1. Get a sample make a histogram and guess a model that fits the histogram.

Step 2. Estimate parameter using MOM (or MLE) estimate. Overlay the histogram from step 1 with a frequency plot for the model distribution with the estimated parameter.

Step 3. Use model with estimated parameter to generate sampling distribution and find its SD.

Discuss with a neighbor the following code.

Note we generally don't know the SD of the ML estimator. In this simple example $SD(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{\lambda}{\sqrt{n}} = \frac{1}{\sqrt{n}}$. For this reason we need to bootstrap the SE even if we assume our distribution is Pois(1).

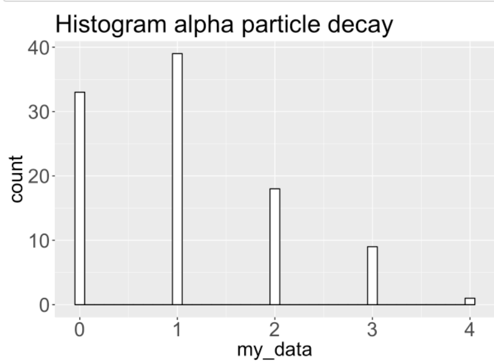
Step 1: visualize data

```
#pop is Pois(1)
sample_size <- 100
B <- 1000 #number samples

my_data <- rpois(sample_size,lambda=1) #my sample
df.my_data<- data.frame(my_data) #make data frame
head(df.my_data)
```

```
##   my_data
## 1       1
## 2       2
## 3       1
## 4       0
## 5       1
## 6       1
```

```
#make histogram
df.my_data %>% ggplot(aes(x=my_data)) +
  geom_histogram(binwidth = .1,col="black",fill="white") +
  labs(title="Histogram alpha particle decay",x="my_data",y="count") +
  theme(
    axis.text = element_text(size = 20),
    plot.title = element_text(size = 25),
    axis.title=element_text(size=20))
```



Based shape of distribution we decide to model as Pois(λ).

Step 2: model data as probability distribution with MLE (or MOM) parameter estimate

We saw in class MLE predicts: $\hat{\lambda} = \bar{X}$.

```
#We find estimated parameters
lambda_hat <- mean(my_data)
lambda_hat
```

```
## [1] 1.06
```

We check for fit of our model by generating random numbers with parameter lambda_hat.

```
theory <- rpois(sample_size, lambda=lambda_hat)
data.df <- data.frame(my_data,theory)
head(data.df)
```

```
##   my_data theory
## 1       1      0
## 2       2      2
## 3       1      1
## 4       0      0
## 5       1      0
## 6       1      2
```

```r
#make histogram
data.df %>% ggplot() +
  geom_histogram(aes(x=data.df$my_data), binwidth = .1,col="black",fill="white") +
  geom_freqpoly(aes(x=data.df$theory),col="red",show.legend = TRUE) +
  labs(title="Theory versus data",x="my_data",y="count") +
  theme(
    axis.text = element_text(size = 20),
    plot.title = element_text(size = 20),
    axis.title=element_text(size=15))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Step 3: Find SE

```r
find_lambda <- function(sample_size){
data <- rpois(sample_size, lambda=lambda_hat)
mean(data)
}

lambda_vector <-  replicate(B, find_lambda(sample_size))
sd(lambda_vector)
```

```
## [1] 0.1056974
```

```r
#expect 1/sqrt(sample_size)
```

1b) The nonparametric bootstrap of estimator SE

In the nonparametric bootstrap we find a sample then we resample that sample many times.

Discuss with a neighbor the following code.

```
sample_size <- 100
B=1000
my_data <- rpois(sample_size,lambda=1)  #my sample
head(my_data)
```

```
## [1] 2 0 0 0 0 0
```

```
#we will run this function many times
find_lambda_np <- function(){
  resample <- my_data %>% sample(replace=TRUE)
  mean(resample)
}

lambda_vec <-  replicate(B, find_lambda_np())
se.lambda=sd(lambda_vec)
se.lambda
```

```
## [1] 0.1040743
```

```
#expect 1/sqrt(sample_size)
```

To experiment cut and paste the above code in the console and run.

```
library(ggplot2)
library(dplyr)
```

2) 95% CI of parameters of λ of Pois(λ) distribution using parametric bootstrap

Next we find the $(1-\alpha)100\,\%$ CI for alpha=.05 (i.e. 95% CI)

Discuss with a neighbor the following code where we calculate the CI of λ.

```
alpha=.05
CI.lambda= 2*lambda_hat -quantile(lambda_vec,c(1-alpha/2,alpha/2))
as.vector(CI.lambda)
```

```
## [1] 0.99 1.41
```

Sec 8.5.2 Large sample theory for MLE

Thm - P277 Rice

The MLE $\hat{\Theta}_{ML}$ for Θ is asymptotically (i.e. for large sample size n) unbiased and normal.

More precisely, for large n,

$$\hat{\Theta}_{ML} \approx N\left(\Theta_0, \frac{1}{nI(\Theta_0)}\right)$$ where

$I(\Theta_0)$ is the Fisher info at the value Θ_0 of Θ.

Fischer info (FI)

What does FI measure?

Lets say a RV X and param Θ are related with some known density function $f(x|\Theta)$

The FI answers the question:

How useful is the RV X in determining Θ.

ex lets look at $\log f(x|\Theta)$ for two circumstances where x is really informative and really uninformative about Θ.

informative $X \sim N(\mu, 1)$

$$\ell(\mu) = \log f(x|\mu)$$

$$= \log\left(\frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2}\right)$$

$$= \boxed{\log\left(\frac{1}{\sqrt{2\pi}}\right) - (x-\mu)^2}$$

uninformative

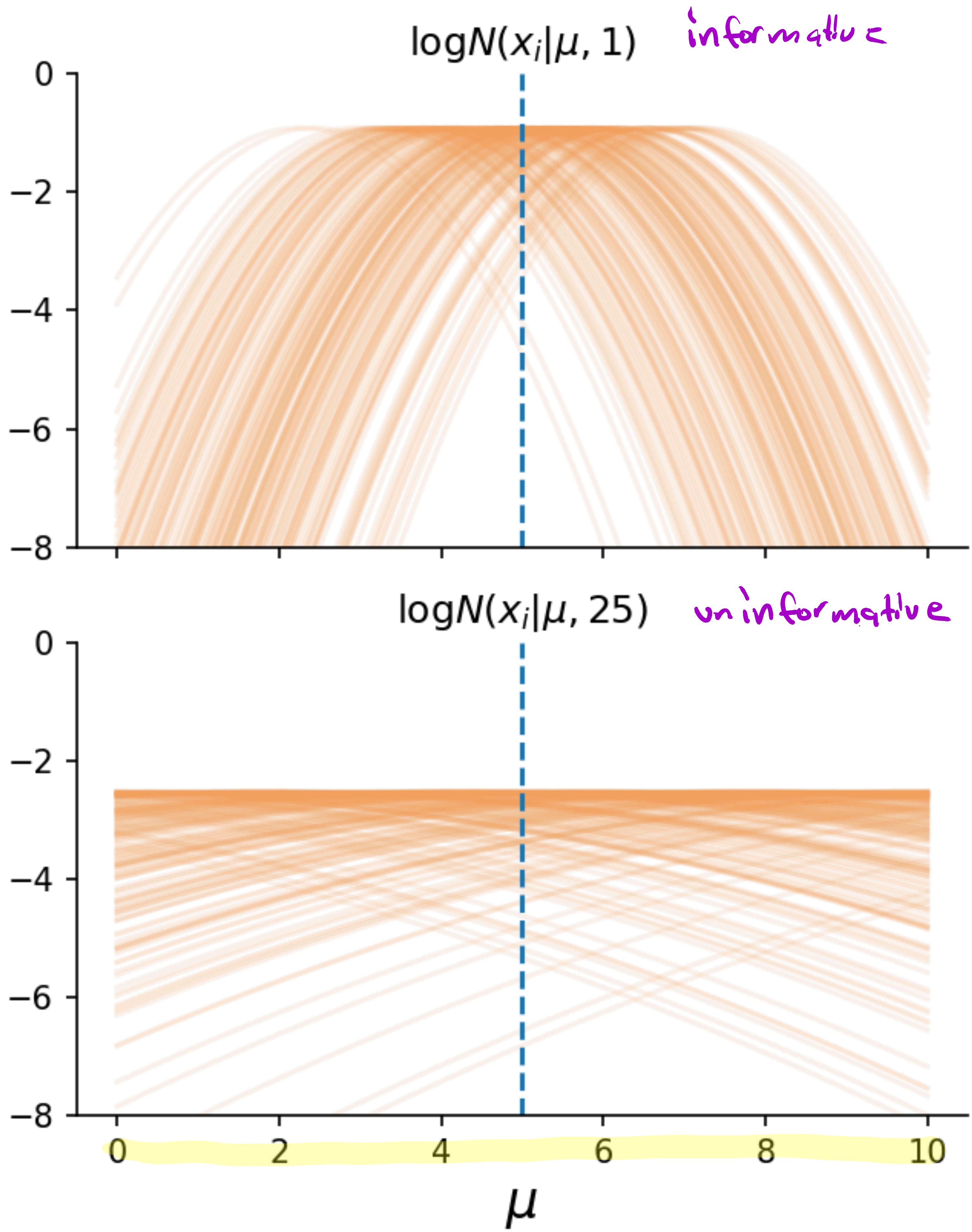
$X \sim N(\mu, 25)$

$$\boxed{\ell(\mu) = \log\left(\frac{1}{\sqrt{2\pi \cdot 25}}\right) - \left(\frac{x-\mu}{5}\right)^2}$$

we plot $\ell(\mu)$ for 1000 individual x

for both $N(\mu,1)$ and $N(\mu,25)$.

Suppose $\mu = 5$.



$\log N(x_i|\mu, 1)$
informative
0
−2
−4
−6
−8
$\log N(x_i|\mu, 25)$
uninformative
0
−2
−4
−6
−8
0
2
4
6
8
10
μ

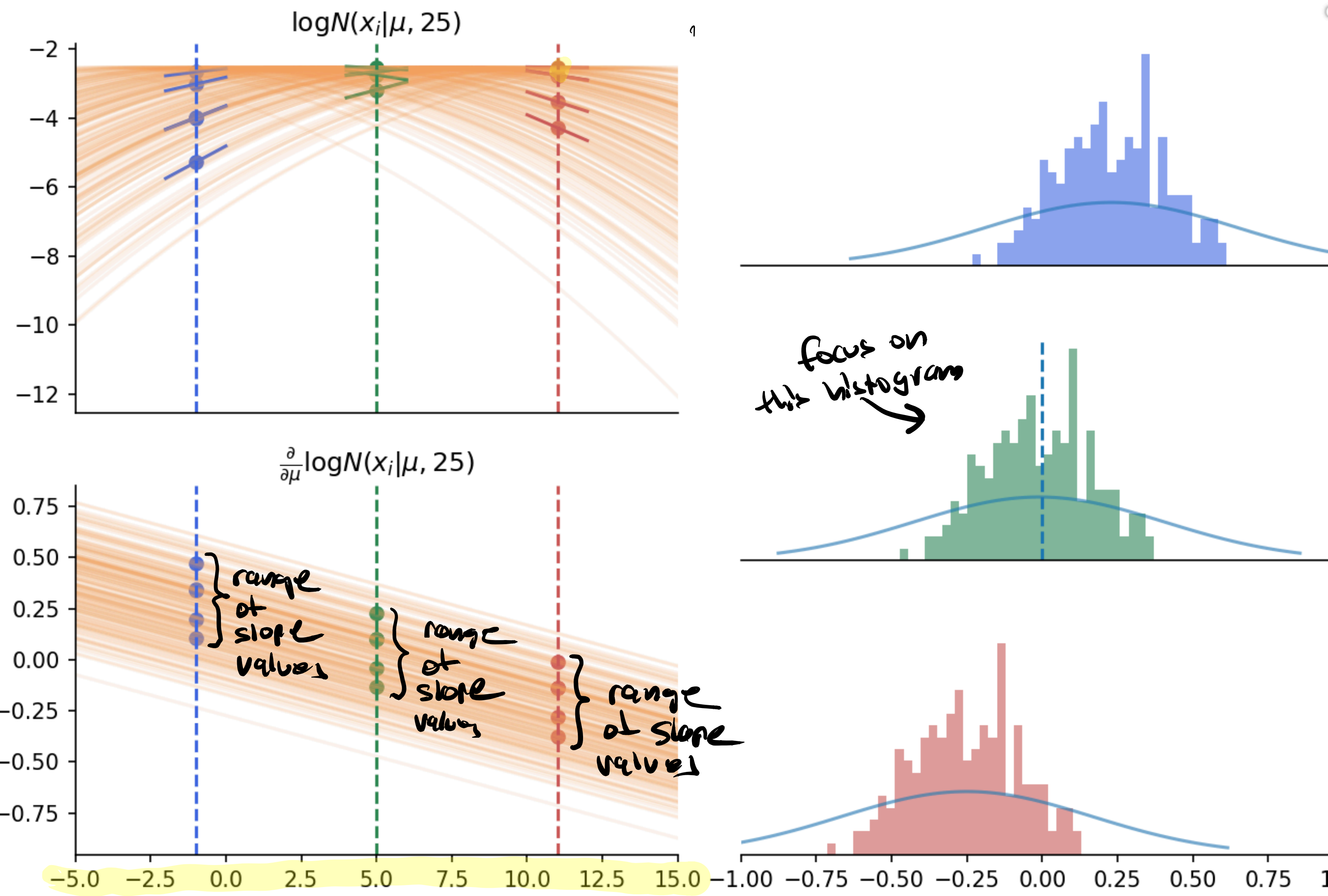
Each curve is providing their own vote of the true parameter location (peak).

How do we measure that a curve has a tighter peak?

It is done by looking at the slopes of the curves (called the <u>score function</u>)

$$\text{Score}_x(\theta) = \frac{\partial}{\partial \theta} \log(f(x|\theta))$$

we pick a few values of μ and look to evaluate the score function and make a histogram



$\log N(x_i|\mu, 25)$
$\frac{\partial}{\partial \mu}\log N(x_i|\mu, 25)$
range of slope values
range of slope values
range of slope values
focus on this histogram
−2
−4
−6
−8
−10
−12
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−5.0 −2.5 0.0 2.5 5.0 7.5 10.0 12.5 15.0
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75

Notice

① We will concentrate on the middle (green) histogram of scores. The above pictures is only for $N(5,25)$. If we make the pictures for $N(5,1)$ the middle histogram would have been wider since there would have been very steep uphill slopes and steep downhill slopes.

② The middle (green histogram) is centered at zero. This is true for both $N(5,1)$ and $N(5,25)$.

$$\underline{\text{i.e}} \quad E\left(\frac{\partial}{\partial\theta_0} \log f(x|\theta)\Big|_{\theta=\theta_0}\right) = 0$$

$\theta_0 = 5$

So the spread of the score at $\theta=\theta_0$ is a measure of how informative x is in estimating θ_0.

Large spread of the histogram of the score at $\theta=\theta_0$ is <u>more</u> informative.

X has a large FI if the histogram of of the score at $\theta=\theta_0$ has big variance.

<u>Def^n</u> FI is the variance of the middle histogram above.

$$I(\theta_0) = \operatorname{var}\left(\frac{\partial}{\partial\theta} \log f(x|\theta)\Big|_{\theta=\theta_0}\right)$$

Fisher info $\quad$ 5