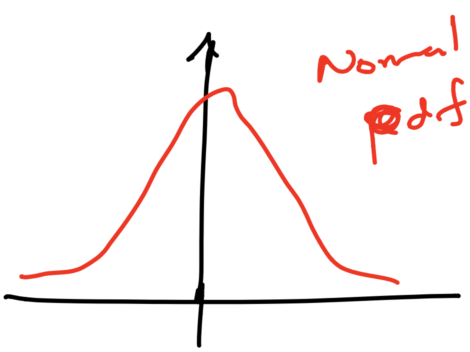
Permutation & Randomization

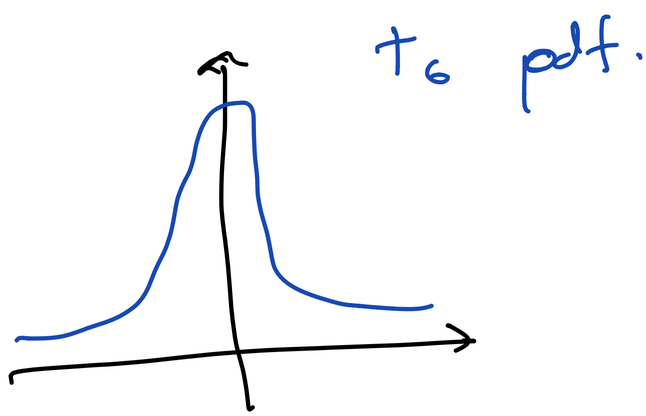
- Recall that last week we conducted a hyp. test on the activity of gene 136 on 2 groups of Leukemia Patients.

Some had ALL diagnosis & some had an AML diagnosis.

- We computed a 2-sample t-test to assess the significance of the effect of ALL diag. vs. AML diag.

- The central assumption of a t-test is that the t-test statistics come from a Normal population.





- Fisher suggests the use of randomization to avoid this assumption.

Not based

on diagnosis

⇒ Taking random groups from the data that are of the same size as the tested groups.

(in our case $n_1 = 47$, $n_2 = 25$)

Then, compute t-statistic for each randomly sampled pair of groups & get a histogram.

→ Let's do it in the computer!

The Cramér-Rao Bound

To finish our discussion about the properties of MLE in a 1-dim setting, suppose that $\tilde{\theta} = t(\underline{x})$ is any unbiased estimate of θ based on iid sample $\underline{X} = (X_1, \dots, X_n)$ from $f_\theta(\underline{x})$. That is:

$$\theta = E_\theta\{t(\underline{x})\}$$

Then, the Cramér-Rao lower bound states that the variance of $\tilde{\theta}$ exceeds Fisher Information Bound $\frac{1}{nI_\theta}$

$$\boxed{\mathrm{Var}_\theta\{\tilde{\theta}\} \geq \frac{1}{nI_\theta}}$$

Cramér-Rao Bound.

- MLE has variance at least as small as the best unbiased estimate of θ.
- MLE is not unbiased in general, but its bias decreases $\underbrace{O\left(\frac{1}{n}\right)}_{\text{Order}}$, making the comparison w/ unbiased estimates adequate.

Conditional Inference

- Say we have iid sample

$$X_i \overset{iid}{\sim} N(\theta, 1)$$

which has produced estimate $\hat{\theta} = \overline{X}$.

However, people conducting the sampling initially disagreed on what the sample size would be, so they flipped a coin to decide:

$$n = \begin{cases} 25 & \text{prob } 1/2 \\ 100 & \text{prob } 1/2 \end{cases}$$

$n = 25$ won.

Q: what is $\hat{\sigma}_{\bar{x}}$ according to a classic frequentist?

$$\left[\frac{1}{2} \cdot \frac{\overbrace{\sigma^2}^{=1}}{25} + \frac{1}{2} \cdot \frac{\overbrace{\sigma^2}^{=1}}{100}\right]^{1/2} = 0.158$$

However, using conditional inference we just care about what really happened.

$$\left[\frac{\sigma^2}{25}\right]^{1/2} = 0.2.$$

Fisher's Arguments for conditional inference:

1. More relevant inferences (only determined by what really happened)
2. Simpler inferences (we didn't need to assess any relationship between the result and the sample size selection step).

Example : Using **Observed** Fisher Information instead of Fisher Inf (which is an expected value).

Rather than using $\hat{\theta} \overset{a}{\sim} N(\theta, 1/(n I_\theta))$

Fisher suggested using $\hat{\theta} \overset{a}{\sim} N(\theta, 1/I(\underline{x}))$

where $I(\underline{x})$ is the observed Fisher Information.

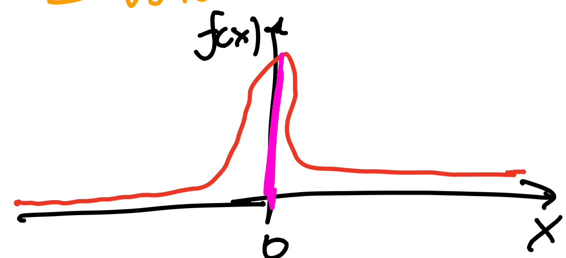
$$I(\underline{x}) \triangleq -\ddot{\ell}_x(\hat{\theta}^{MLE}) = -\frac{\partial^2}{\partial \theta^2} \ell_{\underline{x}}(\theta) \Big|_{\hat{\theta}^{MLE}}$$

Of course $E[I(\underline{x})] = n \cdot I_\theta$, so in large samples the observed Fisher Information is same as Fisher Information. But in smaller samples Fisher suggested $I(\underline{x})$ gave a better idea of $\hat{\theta}$'s accuracy.

Cauchy Distribution

- Distribution of ratio of 2 indep Normals
- Exp. value is undefined
- Variance undefined.

$$f_\theta(x) = \frac{1}{\pi} \cdot \frac{1}{1 + (x-\theta)^2}$$



- Take $n = 20$ samples from it, repeatedly (10000 times).

- we'll get $1/I(\underline{x})$ for each sample and see how much they vary.
- Note that $\hat{\Theta}^{MCE} = \Theta$

---

(1) Sampled vectors of size 20 from a Cauchy distr for 10000 replicates.

(2) For each replicate we got the $\boxed{\hat{\Theta}^{MCE}}$ by minimizing $\ell_x(\Theta)$.

(3) We saved the values of $\hat{\Theta}^{MCE}_{(b)}$ for each replicate $b = 1, \dots, 10000$

(4) We substituted $\hat{\Theta}^{MCE}$ into $-\ddot{\ell}_x(\cdot)$ to get $I(\underline{x})$

(5) We computed $1/I(\underline{x})$ $\hat{\sigma}^{\hat{\Theta}^{MCE}}$ for each replicate and saved it.

→ we're are saving both $\hat{\Theta}^{MCE}$ and our estimators for their variances.

(6) We grouped the replicates with

small, medium, larger, large values
for $\hat{\sigma}_{\hat{\theta}}^{MCE}$.

(7) We computed the variance (using the sample variance formula) for $\hat{\theta}$ MCE's for each group of $\hat{\sigma}_{\hat{\theta}}^{MCE}$ and noticed the getting $\hat{\sigma}_{\hat{\theta}}^{MCE}$ using $1/I(\underline{x})$ is <u>much</u> closer to the sample variance than $\frac{1}{I_\theta}$. <u>Mic Drop.</u>