hypothesis testing

- Let $X_1, ...X_n$ ~ $p(x;\theta)$. the null hypothesis H_0 is the hypothesis that $\theta = \theta_0$ (the hypothesis that it is some fixed value). The opposite of this is $\theta \neq \theta_0$. More generally, H_0 is $\theta \; \epsilon \; \theta_0$ versus $\theta \; \epsilon \; \theta_1$ where $\theta_0 \cap \theta_1 = \emptyset$. The question is whether H_0 is true or false.
- Recall this table:

	Decision	
	Retain H_0	Reject H_0
H_0 true	$\surd$	Type I error (false positive)
H_1 true	Type II error (false negative)	$\surd$

- Hypothesis testing involves:
  - choosing a test statistic T that is some function of the data
  - choose a rejection region R
  - if T is in R we reject H_0, other we retain H_0

  Really, the rejection region is some subset of the sample space.
- in hypothesis testing we'd like the probability of false positive to be below some small number α. This number is picked depends on the context/use case and how much we are willing to tolerate for that use case.

  This means that the probability of rejecting the null hypothesis shouldn't be greater than α if it is true that $\theta = \theta_0$ (or $\theta \; \epsilon \; \theta_0$). we'd like to make it so that the probability is high otherwise. Larry depicts this graphically as:

  where the straight line is $p(rejecting) = \alpha$. this $p(rejecting)$ is called the power function. note that the probability of rejecting can be less than α if the null is true (this is a good thing). α is an imposed upper bound on rejecting if the null hypothesis is true.

- A test is size α if the probability of rejecting the hypothesis when it is correct is $\leq \alpha$.
- It can be shown that that the test should be $\frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} > z_\alpha$ where z_α is such that $P(Z > z) = \alpha$ where Z ~ $N(0, 1)$. proof in Example 3.
- Neyman Pearson test:
  - In the case where H_0: $\theta = \theta_0$ and H_1: $\theta = \theta_1$.
    The test statistic is $T_n = \frac{L(\theta_1)}{L(\theta_0)} = \frac{p(X_1 ... X_n; \theta_1)}{p(X_1 ... X_n; \theta_0)}$. now the test is $T_n > c$. we choose c so that we $p(rejecting) = \alpha$ (test size is α).
  - this is neat but not super useful since it is not often we have this kind of hypothesis setup.
- Wald test:
  - $H_0 : \theta = \theta_0$, $H_1 : \theta \neq \theta_0$.
  - We have an estimator that is asymptotically normal $\hat{\theta}$, typically MLE.
  - $T_n = \frac{\hat{\theta} - \theta_0}{\hat{se}}$.
  - If H_1 is true, we've seen that T_n ~ $N(0, 1)$. We reject if $|T_n| > z_{\alpha/2}$ where the test is size α. The dividing by 2 is because in a normal centered at 0, we have to take into account that it could be too big or too small. We want to reject both.
  - Note that this rejection strategy works when H_0 is true (which is exactly the condition under which we want it to work)
- Likelihood ratio test:
  - this can be used for vectors and not just scalars.

This test is simple: reject H_0 if $\lambda(x_1, \ldots, x_n) \leq c$ where

$$\lambda(x_1, \ldots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} = \frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}$$

where $\widehat{\theta}_0$ maximizes $L(\theta)$ subject to $\theta \in \Theta_0$.

  - Sometimes, (depending on the model), this reduces to the wald test. in general it does not
  - we will work with $-2\, log(\lambda)$ instead of with λ. in general $-2\, log(\lambda)$ converges in distribution to chi-square, under a specific difference in degrees of freedom. proof is on page 7. so using that fact, we determine what c should be so that the probability of type I error stays within α.
- When you design a test and pick some data, the p-value is the smallest α at which we would reject H_0.
- In other words, we reject the null hypothesis when our p-value for a sample is less than α.
- Permutation test:
  - $X_1, ... X_n$ ~ F and $Y_1, ... Y_m$ ~ G and the hypothesis is that $H_0 : F = G$.
  - Now let $Z = (X_1, ... X_n, Y_1, ... Y_m)$. Pick a (any!) test statistic and compute it. Now do a random permutation of Z and re-compute the test statistic. Do this a large of times N. Now consider

$$\frac{1}{K}\sum_{j=1}^{K} I(T^{(j)} > T).$$

Under the null hypothesis this has a uniform distribution (think about why). Now, we can pick an α and either reject or retain.