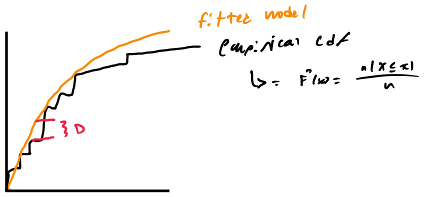
2.5.1 → Kolmogorov Smirnov test

→ overview → tests similarity in cumulative probabilities for the proposed distribution

→ Hypotheses → H_0: proposed distribution adequately fits the data
H_1: — — — does not — — - - - - -



→ Test statistic → conceptually takes the max difference between the empirical cdf + fitted (proposed) cdf

$$\boxed{D_j = \max\left[\,\left|\hat{F}(x_{(j)}) - \overset{\text{fitted cdf}}{F^*}(x_{(j)})\right|,\ \left|\hat{F}(x_{(j-1)}) - F^*(x_{(j-1)})\right|\,\right]}$$

→ run through each pair + take largest

$\Rightarrow D = \max_j D_j$

→ take largest difference

→ critical regions will be given →

α	cv
0.10	$1.22/\sqrt{n}$
0.05	$1.36/\sqrt{n}$
0.01	$1.63/\sqrt{n}$

2.5.2 → Chi-square GoF test

→ overview → conceptually, we discretize the sample space
⇒ then do test counts in each interval

→ Hypotheses → H_0: fits data well
H_1: does not fit data well

→ T.S. → $\boxed{\chi^2 = \sum_{j=1}^{k}\left(\frac{o_j - e_j}{\sqrt{e_j}}\right)^2 \sim \chi^2_{k-1-\ell}}$

↳ # of additional parameters that need estimated from test distribution

→ o_j = observed count
e_j = expected count
↓ = $P(x \in \text{interval}_j \mid \theta_0) * n$

2.5.3 → Chi-square test of ⫫

→ overview → for contingency table data, testing the ⫫ of two variables

	b_1	b_2	
a_1	n_{11}	n_{12}	Σ→
a_2	n_{21}	n_{22}	Σ→
a_3	n_{31}	n_{32}	Σ→
	Σ↓	Σ↓	n

(B across columns, A down rows)

→ Hypotheses → H_0: ⫫
H_1: not ⫫

→ Test statistic → $\chi^2 = \sum_{a=1}^{A}\sum_{b=1}^{B}\frac{(\overset{\text{observed count}}{o_{ab}} - e_{ab})^2}{e_{ab}} \sim \chi^2_{(A-1)(B-1)}$

↳ expected count = $\frac{\text{row total} \times \text{column total}}{n}$

2.5.4 → Likelihood ratio test

→ overview → test whether models come from different distributions

→ Hypotheses → H_0: Data from distribution A
H_1: — — — — — — B

(e.g. $\alpha = 6$, $\alpha \ne 6$ { $\text{Exp}(\beta)$, $\text{gamma}(\alpha, \beta)$ } $\mu = \mu_0$, $\mu \ne \mu_1$)

model A must be a simpler version of model B (i.e. nested)

→ T.S. → $\lambda = -2\ln\left[\frac{L_0}{L_1}\right]$
$= -2[\ln(L_0) - \ln(L_1)]$
$= -2(\ell_0 - \ell_1) \sim \chi^2_{df_0 - df_1}$

↳ essentially difference in # of free parameters