1) Prob. student i guesses correctly but Not psychic?

$$Pr(h_i \text{ Correct} \mid h_i \text{ Not psychic}) = \underbrace{(1-p)\times\cdots\times(1-p)}_{m} = (1-p)^m$$

$b_1 \ldots b_m$
$(1-p) \times \quad \times (1-p)$

2) Prob. at least one correct student without being psychic?

$$Pr(h_1 \text{ correct} \vee h_2 \text{ correct} \vee \cdots \vee h_{|H|} \text{ correct} \mid \text{Nobody psychic}) =$$

$$= 1 - Pr(h_1 \text{ Not correct} \wedge h_2 \text{ Not correct} \wedge \cdots \wedge h_{|H|} \text{ Not Correct} \mid \text{Nobody psychic})$$

$$= 1 - Pr(h_1 \text{ Not correct} \mid \text{Nobody psychic}) \times Pr(h_2 \text{ Not correct} \mid \text{Nobody psychic})$$

$$\times \cdots \times \underbrace{Pr(h_{|H|} \text{ Not Correct} \mid \text{Nobody psychic})}_{1-(1-p)^m}$$

$$= 1 - \left(1 - (1-p)^m\right)^{|H|}$$

$m = 6$	$\lvert H\rvert = 600$	$p = 1/2$	0.9999921
$m = 6$	$\lvert H\rvert = 466$	$p = 1/2$	? 0.99
$m = 6$	$\lvert H\rvert = 50$	$p = 1/2$	0.54

δ

3) How long should m? to have confidence.

$$1 - \left(1 - (1-p)^m\right)^{|H|} < \delta \qquad \delta = 0.05$$

$$(1-\delta)^{1/|H|} < \left(1 - (1-p)^m\right)$$

$$(1-p)^m < 1 - (1-\delta)^{1/|H|}$$

$$m > \log_{(1-p)}\left(1 - (1-\delta)^{1/|H|}\right)$$

$\delta = 0.05$ (95% confidence)

$|H| = 600 \quad p = 1/2 \qquad m \gtrsim 13.6$

$|H| = 500 \quad p = 1/2 \qquad m \geq 13.25$

$|H| = 50 \quad p = 1/2 \qquad m \approx 9.9$

1) $\boxed{\Pr\left(err_S(h)=0 \mid err_P(h)>\varepsilon\right) = ?}$ $\quad \Pr\left(err_S(h)=0 \wedge err_P(h)>\varepsilon\right) \le \exp(-\varepsilon m)$

$x_1, x_2, \ldots, x_m$; $y_1, y_2, \ldots, y_m \sim P$ made a mistake?

$\Pr_{(x_i,y_i)\sim P}\left(h(x_i)\neq y_i\right) \quad err_P(h) = \Pr\left(h \text{ incorrect on } x_i\right)$

$\Pr\left(err_S(h)=0 \mid err_P(h)>\varepsilon\right)\cdot \underbrace{\Pr\left(err_P(h)>\varepsilon\right)}_{\le 1}$

$\le \Pr\left(err_S(h)=0 \mid err_P(h)>\varepsilon\right)$

So $\Pr\left[err_S(h)=0 \mid err_P(h)>\varepsilon\right]$

$= \Pr\left[h \text{ Correct on } x_1 \wedge \ldots \wedge h \text{ Correct on } x_m \mid err_P(h)>\varepsilon\right]$

(independence)

$= \underbrace{\Pr\left[h \text{ Correct on } x_1 \mid err_P(h)>\varepsilon\right]}_{1-err_P(h)} \times \cdots \times \Pr\left[h \text{ correct on } x_m \mid err_P(h)>\varepsilon\right]$

$\le (1-\varepsilon)^m \underset{\text{Unnamed lemma.}}{\le} \exp(-\varepsilon m)$

2) h_S perfect on S chosen after seeing S.

$\Pr\left(err_S(h_S)=0 \mid err_P(h_S)>\varepsilon\right) = ? \quad \le (1-\varepsilon)^m$

Nope! Incorrect $\quad \Pr\left[h_S \text{ is correct } x_i\right] = 1$, $\le (1-\varepsilon)$ ✗

Correct way of doing (2): Want to give gaurantee on $h_S \in H$, instead I give a gaurantee on ALL $h \in H$.

Previously for a fixed $h \in H$: $\Pr\left(err_S(h)=0 \wedge err_P(h)>\varepsilon\right) \le \exp(-\varepsilon m)$.

Previously for a fixed $h \in \mathcal{H}$: $\Pr\left(err_S(h) = 0 \wedge err_P(h) > \varepsilon \right) \leq \exp(-\varepsilon m)$

is small

$$\Pr\left(err_S(h_S) = 0 \wedge err_P(h_S) > \varepsilon \right)$$

$$\leq \Pr\left[\exists h \in \mathcal{H} \quad err_S(h) = 0 \text{ and } err_P(h) > \varepsilon \right]$$

$$\leq \Pr\left[\left(err_S(h_1) = 0 \text{ and } err_P(h_1) > \varepsilon \right) \vee \dots \vee \left(err_S(h_{|H|}) = 0 \text{ and } err_P(h_{|H|}) > \varepsilon \right) \right]$$

Union bound.

$$\leq \underbrace{\sum_{h \in \mathcal{H}}} \underbrace{\Pr\left[err_S(h) = 0 \text{ and } err_P(h) > \varepsilon \right]}_{\leq \exp(-\varepsilon m)}$$

$$\leq |H| \cdot \exp(-\varepsilon m)$$

Importance?

1 - Upper bound on the prob. of "bad event".

"Bad event": $\exists$ h that's good on samples but bad on the distribution.

2) Independent of P, X,

Does depend on $|H|$.

| ... bound gets better.

Does depend

3) $m \uparrow \qquad |H| \exp(-\varepsilon m) \downarrow$ upperbound gets better.

How large m should be so $H \exp(-\varepsilon m) \leq \delta$ $\leftarrow 0.05$

$$\underline{\underline{|H|}} \cdot \exp(-\varepsilon m) \leq \underline{\underline{\delta}} \quad \xrightarrow{\exp}$$

$$\exp\left(\ln(|H|) - \varepsilon m\right) \leq \exp(\ln(\delta))$$

$$\ln(|H|) - \varepsilon m \leq \ln(\delta) = -\ln\left(\frac{1}{\delta}\right)$$

$$\frac{1}{\varepsilon}\left(\ln(|H|) + \ln\left(\frac{1}{\delta}\right)\right) \leq \underline{\underline{m}}$$

# of samples sufficient for learning, as long as, there is one $h \in H$ that's consistent with $\underline{S}$.