Lemma 4.1. *Let P be any distribution and consider a selection function g_θ with a threshold θ whose coverage is $c(\theta, P)$. Let $0 < \delta < 1$ be given and let $\hat{c}(\theta, S_m)$ be the empirical coverage w.r.t. the set S_m, sampled i.i.d. from P. Let $b^*(m, m \cdot \hat{c}(\theta, S_m), \delta)$ be the solution of the following equation:*

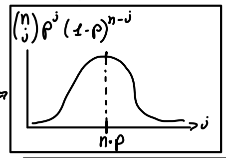
$$arg\min_b \left(\sum_{j=0}^{m \cdot \hat{c}(\theta, S_m)} \binom{m}{j} b^j (1-b)^{m-j} \leq 1-\delta \right).$$

Then,

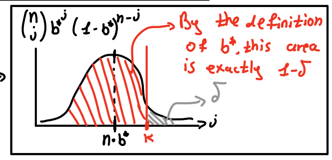
$$\boldsymbol{Pr}_{S_m}\{c(\theta, P) < b^*(m, m \cdot \hat{c}(\theta, S_m), \delta)\} < \delta.$$

Proof. We aim to establish an upper bound, b^*, for the coverage, $c(\theta, P)$, that holds with a probability of at most δ. The coverage $c(\theta, P)$ is equivalent to the probability that $g_\theta(x) = 1$ for a sample $x \sim P$, we will define this event as a 'success'. Define:

$$\mathrm{Bin}(n, k, p) \triangleq \sum_{j=0}^{k} \boxed{\binom{n}{j} p^j (1-p)^{n-j}}, \tag{1}$$

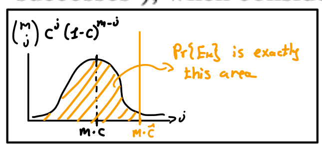


$$\boxed{\begin{aligned} b^* &\triangleq \overline{\mathrm{Bin}}(n, k, \delta) \\ &\triangleq \arg\min_b \left(\mathrm{Bin}(n, k, b) \leq 1-\delta\right). \end{aligned}} \tag{2}$$



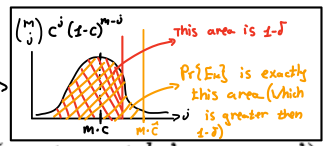
Let E_k be the event that at most $k = m \cdot \hat{c}(\theta, S_m)$ samples x satisfy $g_\theta(x) = 1$ (or at most k 'successes'), when considering $c(\theta, P)$ to be the probability of a single success. Thus,

$$\boxed{Pr\{E_k\} = \mathrm{Bin}(m, m \cdot \hat{c}(\theta, S_m), c(\theta, P)).} \xrightarrow[c \equiv c(\theta, P)]{\hat{c} \equiv \hat{c}(\theta, S_m)} \tag{3}$$



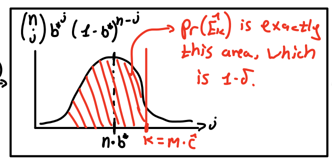
By definition, there is a probability of at most δ, that $k = m \cdot \hat{c}(\theta, S_m)$ 'successes' would fall into the right tail of size δ of the binomial (where the probability of a single success is $c(\theta, P)$). If this happens, then we get that the probability for at most $k = m \cdot \hat{c}(\theta, S_m)$ 'successes' is greater than $1 - \delta$. Mathematically,

$$\boxed{Pr\{E_k\} > 1 - \delta.} \tag{4}$$



Let $\hat{E}_k$ be the event that at most $k = m \cdot \hat{c}(\theta, S_m)$ samples x satisfy $g_\theta(x) = 1$ (or at most k 'successes'), when the probability for a single success is b^*. It holds $\boxed{2}$ that,

$$\boxed{Pr\{\hat{E}_k\} = \mathrm{Bin}(m, m \cdot \hat{c}(\theta, S_m), b^*) = 1 - \delta.} \xrightarrow{\hat{c} \equiv \hat{c}(\theta, S_m)} \tag{5}$$

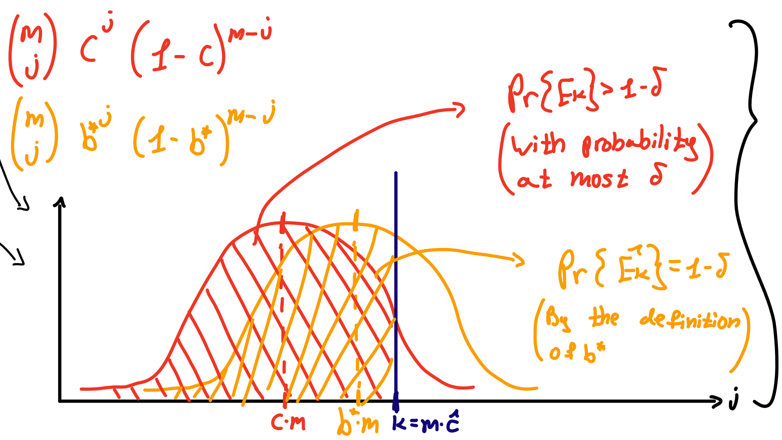


Therefore,

$$Pr\{E_k\} > Pr\{\hat{E}_k\}. \tag{6}$$

Consider the following statement,
let p_1, p_2 be two different probabilities for a single success. If the probability of at most k successes (out of n attempts) is more likely when considering p_1 as the probability of a single success, rather than p_2, we can conclude that $p_1 < p_2$.

Therefore, since the event of E_k occurs with higher probability then the event $\hat{E}_k$, see Eq. $\boxed{6}$, we can conclude that $c(\theta, P) < b^*$. This deduction is only true if we assume that $Pr\{E_k\} > 1 - \delta$, and this assumption is true with probability at most δ. Hence, we get that with probability at most δ, it holds that $c(\theta, P) < b^*$. □



$Pr\{E_k\} > Pr\{\hat{E}_k\} \Rightarrow c < b^*$