such a splitting is referred to as *modular* In our Rényi-Ulam game variation, the expected values of FDR and FOR are the error rates e_1 and e_0 in case the truthful answer being yes and no, respectively. Fig. displays the error rates e_0 and e_1 as functions of θ. For $\theta = 31$, we compute $e_0 \approx 0.052$ and $e_1 \approx 0.007$, i.e. we have an error rate of 0.052 for rejecting and an error rate of 0.007 for confirming a base pair.

0.1. Entropy

To quantify the uncertainty of an ensemble, we define the *structural entropy* of an ensemble, Ω, of an RNA sequence, $\mathbf{x}$, as the Shannon entropy

$$H(\Omega) = -\sum_{s \in \Omega} p(s) \log_2 p(s),$$

Proposition implies that a sample with small structural entropy contains a distinguished structure and a proof is given in . We refer to a sample having a distinguished structure of probability at least λ as being λ-*distinguished*.

Next we quantify the reduction of a bit query on an ensemble. Recall that the associated r.v. $X_{i,j}$ of a base pair (i, j) partitions the sample Ω into two disjoint sub-samples Ω_0 and Ω_1, where $\Omega_k = \{s \in \Omega : X_{i,j}(s) = k\}$ $(k = 0, 1)$. The *conditional entropy*, $H(\Omega|X_{i,j})$, represents the expected value of the entropies of the conditional distributions on Ω, averaged over the conditioning r.v. $X_{i,j}$ and can be computed by

$$H(\Omega|X_{i,j}) = (1 - p_{i,j})H(\Omega_0) + p_{i,j}H(\Omega_1).$$

Then the *entropy reduction* $R(\Omega, X_{i,j})$ of $X_{i,j}$ on Ω is the difference between the *a priori* Shannon entropy $H(\Omega)$ and the conditional entropy $H(\Omega|X_{i,j})$, i.e.

$$R(\Omega, X_{i,j}) = H(\Omega) - H(\Omega|X_{i,j}).$$

$\mathbb{P}(s \in \Omega^*) = (1 - e_0)^{l_0}(1 - e_1)^{l_1}$, where l_0 and l_1 denote the number of No-/Yes-answers to queried base pairs along the path, respectively. Fig. displays the distribution of l_1. We observe that l_1 has a mean around 5, i.e., the probabilities of queried base pairs being confirmed and being rejected are roughly equal. For $l_0 = l_1 = 5$, we have a theoretical estimate $\mathbb{P}(s \in \Omega^*) \approx 0.736$. In Fig. we present that $\mathbb{P}(s \in \Omega^*)$ decreases as the error rate e_0 increases, for fixed $e_1 = 0.01$.

1. Information theory

As the Boltzmann ensemble is a particular type of discrete probability spaces, the information-theoretic results on the ensemble trees will be stated in the more general setup. Let $\Omega = (\mathcal{S}, \mathcal{P}(\mathcal{S}), p)$ be a discrete probability space consisting of the sample space $\mathcal{S}$, its power set $\mathcal{P}(\mathcal{S})$ as the σ-algebra and the probability measure p. The *Shannon entropy* of Ω is given by

$$H(\Omega) = -\sum_{s \in \mathcal{S}} p(s) \log_2 p(s),$$