The house buying problem

I have n houses to choose from, following viewings. All match my requirements for multiple bathrooms, open-plan kitchens, secluded gardens with mega-shed potential, houses that are not to too close to the M27, but within cycling distance of my workplace. However, the market is so competitive that once I reject a choice, I cannot put in an offer later since the house gets immediately snapped up!
What is my optimum strategy to have the maximum chance of choosing the *best house overall*?

The strategy is, as n increases, to ***reject* the first k houses** and **then select the first house which is better than the best of the first** k. The optimum value for k is: $\boxed{k=\frac{n-1}{e}}$

Assume we reject the first k houses, the probability that pick j of the *remaining* n-k houses **is the best overall** is:

$$\boxed{p(n,k,j)=P\left(\text{house } j \text{ of } n\text{-}k \text{ is the best overall}\right)=P\left(\text{house } j \text{ of } n\text{-}k \text{ is the best}\right)\times P\left(\text{the previous } j\text{-}1 \text{ houses are } \textit{not} \text{ better than the best in the first } k\right)}$$

$$\therefore p(n,k,j)=\frac{1}{n}\times\frac{k}{n+j-1}$$

$$P\left(\text{the previous } j\text{-}1 \text{ houses are } \textit{not} \text{ better than the best in the first } k\right)$$

i.e. the *same probability* that the **best in the first** k+j -1 **houses** is *within* the first k, which were rejected.

$$P\left(\text{house } j \text{ of } n\text{-}k \text{ is the best overall}\right)$$

Assume no prior knowledge and a randomized viewing list so every house has the same chance of being the best.

Hence the probability of choosing the best house using this strategy is:

$$q(n,k)=\sum_{j=1}^{n-k}p(n,k,j)=\frac{1}{n}\times\sum_{j=1}^{n-k}\frac{k}{n+j-1}$$
$$=\tfrac{1}{n}\left(1+k\left(\tfrac{1}{k+1}+\tfrac{1}{k+2}+\ldots+\tfrac{1}{n-1}\right)\right)$$

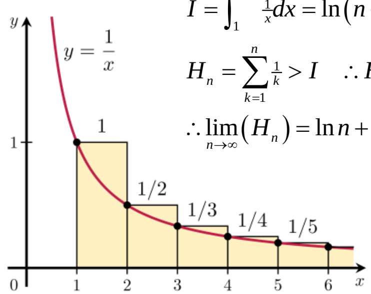
Now: $$\underbrace{1+\tfrac{1}{2}+\tfrac{1}{3}+\ldots+\tfrac{1}{n-1}}_{H_{n-1}}=\underbrace{1+\tfrac{1}{2}+\ldots+\tfrac{1}{k}}_{H_k}+\tfrac{1}{k+1}+\tfrac{1}{k+2}+\ldots+\tfrac{1}{n-1}$$
$$\therefore \tfrac{1}{k+1}+\tfrac{1}{k+2}+\ldots+\tfrac{1}{n-1}=H_{n-1}-H_k$$

$$\boxed{\therefore q(n,k)=\tfrac{1}{n}\left(1+k\left(H_{n-1}-H_k\right)\right)}$$

***Harmonic series* sum**

$$H_n=1+\tfrac{1}{2}+\tfrac{1}{3}+\ldots+\tfrac{1}{n}$$
$$\lim_{n\to\infty}\left(H_n\right)=\ln n+\gamma$$
$$\gamma\approx 0.5772156649$$

Euler–Mascheroni constant



$$I=\int_1^{n+1}\tfrac{1}{x}dx=\ln\left(n+1\right)$$
$$H_n=\sum_{k=1}^{n}\tfrac{1}{k}>I \quad \therefore H_n>\ln\left(n+1\right)$$
$$\therefore \lim_{n\to\infty}\left(H_n\right)=\ln n+\gamma$$

The difference between the harmonic series sum and the integral of $1/x$ tends to a *constant* as n increases.

For large n : $\therefore q(n,k)\approx\tfrac{1}{n}\left(1+k\ln(n-1)-k\ln k\right)=\tfrac{1}{n}+\tfrac{k}{n}\ln\left(\tfrac{n-1}{k}\right)$

$$H_n=1+\tfrac{1}{2}+\tfrac{1}{3}+\ldots+\tfrac{1}{n}\approx\ln n+\gamma$$

$$\frac{\partial}{\partial k}q(n,k)=\tfrac{1}{n}\left\{\ln(n-1)-\ln k-1\right\}$$
$$\therefore\frac{\partial}{\partial k}q(n,k)=0 \quad\text{when}\quad \ln\left(\frac{n-1}{k}\right)=1$$

i.e. what k gives the maximum q?

$$\therefore\frac{n-1}{k}=e\quad\Rightarrow\boxed{k=\frac{n-1}{e}}$$
$$\therefore q_{\max}(n)=q\left(n,k=\frac{n-1}{e}\right)$$
$$=\tfrac{1}{n}+\tfrac{1}{n}\frac{n-1}{e}\ln\left(e\right)$$
$$=\tfrac{1}{n}+\tfrac{1}{e}\left(1-\tfrac{1}{n}\right)$$
$$\rightarrow\boxed{\frac{1}{e}}\quad\text{for large } n$$

$$1/e\approx\tfrac{1}{2.7183}\approx 0.3679$$

so one should reject about 37% of the n houses to choose from before picking the best so far.

Dashed -- lines are the approximate result

$$q\left(n,k\right)=\tfrac{1}{n}+\tfrac{k}{n}\ln\left(\tfrac{n-1}{k}\right)$$

Solid lines with **.** are the exact calculation

$$q\left(n,k\right)=\tfrac{1}{n}\left(1+k\left(H_{n-1}-H_k\right)\right)$$

Probability of best candidate chosen from n candidates after rejecting the first k