Some bounds and approaches useful for (randomized) algorithms

Aravind Srinivasan, with help from student scribes

1 A basic review of continuous distributions

Discrete	Continuous
$X = x$	$X \in [x, x + dx)$
$Pr[X = x]$	$Pr[X \in [x, x + dx)] = f(x)dx$, where $f(x)$ is the "density function" of X
$Pr[a \leq X \leq b]$	$\int_a^b f(x)dx$
$E[X] = \sum_x x \cdot Pr[X = x]$	$E[X] = \int_{-\infty}^{\infty} xf(x)dx$

If follows that if X is always between l, u, then $f(x) \geq 0$ and $\int_l^u f(x)dx = 1$
Note that $Pr[X = x]$ is $f(x)dx = 0 \cdot f(x)$. Therefore the probability of any variable being exactly some value is 0. From this it follows that the probability that two continuous random variables have the same value is 0.

One important continuous distribution is the normal or Gaussian distribution. If the mean is μ and standard deviation is σ, then

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

here.

Another frequently used distribution is the uniform distribution over a bounded range, such as $[0, 1]$. (If this range is $[a, b]$, then the density function $f(x)$ is the *constant* $1/(b-a)$.) The idealized version allows us to pick a real number uniformly at random from a bounded range. In practice we approximate this by a discretization of the space.

2 Convexity and some of its consequences

Recall that a function f is *convex* in the interval $[a, b]$ if for any $[u, v] \subseteq [a, b]$, the graph of f lies below the line segment that joins the points $(u, f(u))$ and $(v, f(v))$: that is,

$$\forall (u, v) \text{ such that } a \leq u \leq v \leq b,\ \forall p \in [0, 1], \quad f(up + v(1 - p)) \leq p \cdot f(u) + (1 - p) \cdot f(v).$$

And, f is called *concave* in $[a, b]$ if the final inequality gets reversed in direction. In case the second derivative f'' exists for $x \in [a, b]$, then f can be shown to be convex in $[a, b]$ iff $f''(x) \geq 0$, and concave in $[a, b]$ iff $f''(x) \leq 0$. Using this, or pictorially, one can verify that:

- $f(x) = x^{2k}$ is convex over the entire real line, for all positive integers k;
- $f(x) = e^{ax}$ is convex over the entire real line, for all reals a;
- $f(x) = \ln x$ and $f(x) = \sqrt{x}$ are concave for the entire range $(0, \infty)$;
- $f(x) = x^3$, $f(x) = x^5$ etc. are concave for $x < 0$, and convex for $x > 0$;
- a linear function is both concave and convex.

The following fact is often useful. Suppose a function f is convex in some domain $D = [a, b]$ and $x_1, \ldots, x_n$ are variables such that $x_i \in D$ and $\sum x_i = c$, then $\sum_i f(x_i)$ is minimized when all x_i are the same (i.e., c/n). Therefore,

$$\sum_i f(x_i) \geq n \cdot f(c/n). \tag{1}$$

Conversely, to *maximize* $\sum_i f(x_i)$ subject to the given constraints, we must push the x_i's as much as possible to their "extreme values" (the definition of this varies from one context to another, but the intuitive meaning remains: "do the opposite of pushing the x_i toward each other".) As can be expected, the situation is exactly reversed for concave functions f: make the x_i equal if we aim to *maximize* $\sum_i f(x_i)$, and do a sort of opposite if the goal is to minimize $\sum_i f(x_i)$

Jensen's Inequality: This is an easy consequence of convexity, and says the following. If f is a convex function in the interval $[a, b]$, then for any random variable X taking values in $[a, b]$,

$$\mathrm{E}[f(X)] \geq f(\mathrm{E}[X]);$$

and, f is a concave function in the interval $[a, b]$, then for any random variable X taking values in $[a, b]$,

$$\mathrm{E}[f(X)] \leq f(\mathrm{E}[X]).$$

We finally note that all the above discussion was focused on uni-variate functions f, but quite a bit of the above also extends to real-valued multi-variate functions $f(x_1, x_2, \ldots, x_n)$.

3 Additional inequalities and some information theory

We begin by examining some additional useful bounds.

- $1 + x \leq e^x \ \forall x$; also, $(1 - \frac{1}{n})^{n-1} \geq \frac{1}{e}$ for $n \geq 2$. In particular, the bound $(1 - x)^t \leq e^{-tx}$ for $x \leq 1$ and $t > 0$ that follows from the first inequality here, is used routinely.
- Stirling's approximation: $F(n) = \sqrt{2\pi n}(\frac{n}{e})^n$, and Robbins' formula for the error of the approximation: $e^{1/(12n+1)} \leq \frac{n!}{F(n)} \leq e^{1/(12n)}$. Note that both $e^{1/(12n+1)}$ and $e^{1/(12n)}$ are very close to 1 even for moderately large n; e.g., $n \geq 5$. Thus, Stirling's formula is a very good approximation for $n!$ if $n \geq 5$, say.
- $(\frac{n}{r})^r \leq \binom{n}{r} \leq (\frac{ne}{r})^r$; in fact, $\sum_{i=0}^{r} \binom{n}{i} \leq (\frac{ne}{r})^r$.
- An alternative bound is as follows. Let $H(\alpha) = -\alpha \log_2 \alpha - (1 - \alpha) \log_2(1 - \alpha)$ be the "binary entropy function" for $0 \leq \alpha \leq 1$; if α is 0 or 1, we define $H(\alpha) = 0$. Then, for $\alpha \leq \frac{1}{2}$,

$$\sum_{i=0}^{\alpha n} \binom{n}{i} \leq 2^{nH(\alpha)}. \tag{2}$$

This last statement, (2), is of particular interest, so we introduce basic information theory next to look at it more closely.

3.1 Basic Information Theory

We first consider the following definitions and observations from information theory. Let X be a random variable which takes on the value a_i with probability p_i for $i = 1 \ldots n$. We loosely define the *entropy* of X as the "amount of randomness in X", given by the function:

$$H(X) = -\sum_{i=1, p_i \neq 0}^{n} p_i \log_2 p_i$$

From this we have the following facts:

- $H(X) \geq 0$, and $H(X) = 0$ iff X is deterministic (one of the p_i equals 1).
- For n-valued X, $H(X)$ is maximized when X uses the uniform distribution, that is,

$$p_1 = p_2 = \ldots = p_n = \frac{1}{n}; \text{ in this case, } H(X) = -\log_2(\frac{1}{n}) = \log_2 n$$

  In particular, $H(X) \leq \log_2 n$; we will see this in the proof of Claim 1.

- For joint distributions, $H(X_1, X_2, \ldots, X_m) \leq \sum_{i=1}^{m} H(X_i)$.

We now consider the following example, which will be useful in proving (2).

Example: $X = \begin{cases} 1 \text{ with probability } p \\ 0 \text{ with probability } 1-p \end{cases}$

From this we have that $H(X) = -p\log_2 p - (1-p)\log_2(1-p)$, which we will call $H(p)$ for simplicity (though it is an abuse of notation since p is a value while X is a random variable). It is easy to see that $H(p)$ is a concave function either by viewing the graph of the function or by taking its second derivative. With this, we can prove (2), which we state next in more generality.

Claim 1 *Let $\alpha \leq \frac{1}{2}$, and let S be any set of n-bit strings such that the average number of 1's in a string in S is $\leq \alpha n$. (That is, the average taken over all strings $s \in S$ of the number of ones in s, is at most αn.) Then $|S| \leq 2^{nH(\alpha)}$.*

Proof Let $X = (X_1, X_2, \ldots, X_n)$ chosen uniformly at random from S. Then $H(X) = \log_2 |S|$. In fact,

$$\log_2 |S| = H(X) \leq \sum_{i=1}^{n} H(X_i)$$

Suppose $\forall i, X_i = \begin{cases} 1 \text{ with probability } p_i \\ 0 \text{ with probability } 1-p_i \end{cases}$; then $H(X_i) = H(p_i)$ as in the example.

Thus $\sum_{i=1}^{n} H(X_i) = \sum_{i=1}^{n} H(p_i)$.

Now notice that we know something about the sum of these p_i's, in particular:

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} E[X_i] = E[\sum_{i=1}^{n} X_i] \leq \alpha n$$

We now use the fact that if we have a sum of concave functions which is equal to a fixed value, that sum is maximized when the functions are equal. Therefore, if $\sum_{i=1}^{n} p_i = t$,

$$\sum_{i=1}^{n} H(p_i) \leq n \cdot H(\frac{t}{n}) \leq nH(\alpha) \text{ since } \alpha \leq \frac{1}{2}$$

Therefore $|S| \leq 2^{nH(\alpha)}$. ■