Notes on the Analysis of Boolean Functions

Ravi Kant
Yahoo Labs, Sunnyvale
rkant@yahoo-inc.com

1. COMPONENTS AND COMPLEMENTS

Let $f : \mathbb{F}_2^n \to \mathbb{R}$ and let its Fourier expansion be

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S(x)$$

Definition For a subset $I \subseteq [n]$ we define the I-*component*, f_I of f as

$$f_I(x) = \sum_{S \subseteq I} \hat{f}_S \chi_S(x)$$

and the I-*complement*, $\bar{f}_I$ as

$$\bar{f}_I(x) = \sum_{S \not\subseteq I} \hat{f}_S \chi_S(x) = f(x) - f_I(x)$$

Remark 1.1.

$$\sum_{x \in \mathbb{F}_2^n} f(x) = \sum_{x \in \mathbb{F}_2^n} f_I(x)$$

since $\hat{f}(\varnothing) = \hat{f}_I(\varnothing)$

We will interpret a vector $x \in \mathbb{F}_2^n$ as the indicator of a subset $S \subseteq [n]$ and (abuse of notation) use the same letter for the vector and the set when the interpretation is clear from the context.

Lemma 1.2. *Let*

$$\delta_I(x) = \begin{cases} 2^{|I|} & \text{if } I \cap x = \varnothing \\ 0 & \text{otherwise} \end{cases}$$

then,

$$f_I = f * \delta$$

Proof.

$$\begin{aligned} f_I(x) &= \sum_{S \subseteq I} \hat{f}_S \chi_S(x) \\ &= \sum_{S \subseteq I} \chi_S(x) \frac{1}{2^n} \sum_{y \in \mathbb{F}_2^n} f(y)\chi_S(y) \\ &= \frac{1}{2^n} \sum_y f(y) \sum_{S \subseteq I} \chi_S(x+y) \\ &= \frac{1}{2^n} \sum_y f(y)\delta_I(x+y) \end{aligned}$$

□

Corollary 1.3. *If* $f : \mathbb{F}_2^n \to \{0,1\}$ *then*

$$\forall x, I,\ 0 \le f_I(x) \le 1$$

Proof. δ_I is a probability density function (i.e. $\delta_I / 2^n$ is a probability distribution.) □

Theorem 1.4. *(Bourgain) Let* $f : \mathbb{F}_2^n \to \{0,1\}$. *Fix a subset* $I \subseteq [n]$. *Then,*

$$2 \sum_{x \in \mathbb{F}_2^n} \bar{f}_I(x)^2 = \sum_{x \in \mathbb{F}_2^n} |\bar{f}_I(x)|$$

Proof.

$$\begin{aligned} LHS &= 2 \sum_{x \in \mathbb{F}_2^n} (f(x) - f_I(x))^2 \\ &= 2f \cdot f + 2f_I \cdot f_I - 4f \cdot f_I \\ &= 2f \cdot f - 2f \cdot f_I \ (\text{since } f_I \cdot f_I = f \cdot f_I) \\ RHS &= \sum_{x \in \mathbb{F}_2^n} |f(x) - f_I(x)| \\ &= \sum_{f(x)=0} f_I(x) + \sum_{f(x)=1} (f(x) - f_I(x)) \ (\text{see corollary}) \\ &= \sum_{x \in \mathbb{F}_2^n} f(x) + \sum_{x \in \mathbb{F}_2^n} (1 - 2f(x)) f_I(x) \\ &= f \cdot f + f \cdot f - 2f \cdot f_I \end{aligned}$$

□

2. LOW DEGREES

Problem Characterize all Boolean functions with degree ≤ 1.

LEMMA 2.1. *The only degree ≤ 1 boolean functions are $\pm\chi_S$ where $|S| \leq 1$.*

PROOF 1. $f = \sum_{i=0}^n \hat{f}_i \chi_i$ where i refers to the empty set when $i = 0$ and the singleton set $\{i\}$ otherwise and let $x^{(i)}$ denote its vector indicator. Since f is boolean,

$$\begin{aligned} f(1) &= \pm 1 = \sum_0^n \hat{f}_i \\ f(x^{(i)}) &= \pm 1 = \hat{f}_0 + \cdots \hat{f}_{i-1} - \hat{f}_i + \hat{f}_{i+1} + \cdots + \hat{f}_n \\ \implies \hat{f}_i &= 0 \text{ or } \pm 1 \end{aligned}$$

Also we know that $\sum_0^n \hat{f}_i^2 = 1$. QED. $\square$

PROOF 1 REHASHED.

$$\begin{aligned} f &= \sum_{i=0}^n \hat{f}_i \chi_i \\ \implies D_i[f] &= \hat{f}_i \chi_0 \text{ (where } 1 \leq i \leq n) \\ \implies \forall x, D_i[f](x) &= \hat{f}_i \\ &= 0 \text{ or } \pm 1 \text{ (since } f \text{ is boolean)} \end{aligned}$$

As in first proof, QED $\square$

3. CORRELATED GAUSSIANS

LEMMA 3.1. *Let $N(\mu, \sigma^2)$ be the normal distribution with mean μ and variance σ^2. Then*

$$N(\mu_1, \sigma_1^2) + N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

where the LHS random variables are assumed to be independent.

LEMMA 3.2. *If $G = H = N(0,1)$ and $E[GH] = \rho$ then $H = \rho G + \sqrt{1-\rho^2}G'$ where $E[G, G'] = 0$*

PROOF. Define $G' = \frac{1}{\sqrt{1-\rho^2}}(H - \rho G)$. Using the preceding lemma, $G' = N(0,1)$. Using $E[GH] = \rho$,

$$\begin{aligned} E[GG'] &= \frac{1}{\sqrt{1-\rho^2}} E(GH - \rho G^2) \\ &= \frac{\rho - \rho \cdot 1}{\sqrt{1-\rho^2}} \\ &= 0 \end{aligned}$$

$\square$

LEMMA 3.3. *Let G,H be standard Gaussians with $E[GH] = \rho$. Then $P[sgn(G) \neq sgn(H)] = \cos^{-1} \frac{\rho}{\pi}$.*

PROOF. Consider the joint distribution $[G, G']$ from previous lemma. We need to compute the relative area of the region $R := \{(x,y) : sgn(x) \neq sgn(\rho x + \sqrt{1-\rho^2}y)\}$. The result follows from the circular symmetry of the joint distribution. $\square$

This leads to the following:

- $W^k(Maj_n) \sim (\frac{2}{k\pi})^{3/2}$, when k is odd, 0 o/w.
- $W^{>=k}(Maj_n) \sim (\frac{2}{\pi})^{3/2} \cdot k^{-1/2}$

4. SOME SPECIAL FUNCTIONS

We will use the following notation: For any subset $S \subseteq [n]$

- e_S and χ_S are functions on the boolean cube
- e'_S and χ'_S are functions on the dual boolean cube.

4.1 Spectra

4.1.1 Neighbourhoods of a character

$$min_n = OR_n = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\widehat{OR}_n = \begin{bmatrix} -1 + \frac{2}{2^n} \\ \frac{2}{2^n} \\ \vdots \\ \frac{2}{2^n} \end{bmatrix} = -e'_\varnothing + \frac{2}{2^n}\chi'_\varnothing$$

Similarly,

$$\widehat{AND}_n = \begin{bmatrix} 1 - \frac{2}{2^n} \\ \vdots \\ \frac{2(-1)^{|S|+1}}{2^n} \\ \vdots \end{bmatrix} = e'_\varnothing - \frac{2}{2^n}\chi'_{[n]}$$

In general,

$$1_S : \mathbb{F}_2^n \to \{-1, 1\}$$

where $1_S(a) = -1$ iff a is the indicator of S, then

$$\hat{1}_S = e'_\varnothing - \frac{2}{2^n}\chi'_S$$

Obvious generalizations:

- $EQ_n = 1 - 2e_\varnothing - 2e_{[n]} \implies \widehat{EQ}_n = e'_\varnothing - \frac{2}{2^n}\chi'_\varnothing - \frac{2}{2^n}\chi'_{[n]}$
- $\widehat{NAE}_n = -e'_\varnothing + \frac{2}{2^n}\chi'_\varnothing + \frac{2}{2^n}\chi'_{[n]}$
- Analogous results for the neighbourhoods of a nontrivial character

4.1.2 Quadratic functions

For $x_1, x_2 \in \mathbb{F}_2^n$ we define

$$IP_2(x_1, x_2) = (-1)^{x_1 \cdot x_2}$$

Clearly:

$$\begin{aligned} \langle IP_2, \chi_{S_1,S_2} \rangle &= \frac{1}{2^{2n}} \sum_{x_1,x_2} (-1)^{x_1 \cdot x_2} (-1)^{x_1 \cdot S_1} (-1)^{x_2 \cdot S_2} \\ &= \frac{1}{2^{2n}} \sum_{x_1} (-1)^{x_1 \cdot S_1} \sum_{x_2} (-1)^{(x_1 + S_2) \cdot x_2} \\ &= \frac{1}{2^{2n}} \sum_{x_1} (-1)^{x_1 \cdot S_1} \cdot 2^n \delta(x_1, S_2) \\ &= \frac{1}{2^n} (-1)^{S_2 \cdot S_1} \end{aligned}$$

Note that this shows IP_2 is self dual.

4.1.3 The Complete Quadratic and Variants

For $x = (x_1, \ldots, x_n) \in \mathbb{F}_2^n$ we define

$$CQ(x) = (-1)^{\sum_{i<j} x_i x_j}$$

$$\begin{aligned}
\langle CQ, \chi_S \rangle &= \frac{1}{2^n} \sum_x (-1)^{S\cdot x} \cdot (-1)^{\binom{|x|}{2}} \\
&= \frac{1}{2^n} [\sum_k (-1)^k (\sum_{\substack{|x|=0(4)\\ S\cdot x=k}} 1 + \\
&\quad \sum_{\substack{|x|=1(4)\\ S\cdot x=k}} 1 - \sum_{\substack{|x|=2(4)\\ S\cdot x=k}} 1 - \sum_{\substack{|x|=3(4)\\ S\cdot x=k}} 1)] \\
&= \frac{1}{2^n} [S_0 + S_1 - S_2 - S_3] \\
&= \hat{f}_S (say)
\end{aligned}$$

The second line in the above equations comes from the fact that $\binom{n}{2}$ is even if n is 0 or 1 mod 4 and odd otherwise.

Now let $\sigma_0(n) = \sum_{i=0(4)} \binom{n}{i}$, similarly $\sigma_k(n) = \sum_{i=k(4)} \binom{n}{i}$. We know that:

$$4\sigma_k(n) = 2^n + 2^{\frac{n}{2}+1} \cos((n-2k)\frac{\pi}{4})$$

Also let $|S| = d$. We now evaluate the above 4 sums in turn:

$$\begin{aligned}
S_0 &= \sum_{S\cdot x=k=0}^{k=d} (-1)^k \sum_{|x|=0(4)} 1 \\
&= \sigma_0(n-d) + (-1)\binom{d}{1}\sigma_3(n-d) + \cdots \\
&\quad +(-1)^k \binom{d}{k} \sigma_{4-k}(n-d) + \cdots \\
&= \sigma_0(n-d)\sigma_0(d) - \sigma_1(n-d)\sigma_3(d) + \\
&\quad \sigma_2(n-d)\sigma_2(d) - \sigma_3(n-d)\sigma_1(d)
\end{aligned}$$

Similarly,

$$\begin{aligned}
S_1 &= -\sigma_0(n-d)\sigma_1(d) + \sigma_1(n-d)\sigma_0(d) + \\
&\quad -\sigma_2(n-d)\sigma_3(d) + \sigma_3(n-d)\sigma_2(d)
\end{aligned}$$

$$\begin{aligned}
S_2 &= \sigma_0(n-d)\sigma_2(d) - \sigma_1(n-d)\sigma_1(d) + \\
&\quad \sigma_2(n-d)\sigma_0(d) - \sigma_3(n-d)\sigma_3(d)
\end{aligned}$$

$$\begin{aligned}
S_3 &= -\sigma_0(n-d)\sigma_3(d) + \sigma_1(n-d)\sigma_2(d) + \\
&\quad -\sigma_2(n-d)\sigma_1(d) + \sigma_3(n-d)\sigma_0(d)
\end{aligned}$$

Substituting for the σ_i

$$\hat{f}_S = \sqrt{\frac{2}{2^n}} \sin((n-2d+1)\frac{\pi}{4})$$

For even n the sin term is $\pm\frac{1}{\sqrt{2}}$ so CQ is bent.

REMARK 4.1. *Since CQ and $\widehat{CQ}$ are symmetric functions we can talk of $CQ(d)$ and $\widehat{CQ}(d)$ where d is the degree of the corresponding inputs. When n is even we have*

$$sgn(\widehat{CQ}(d)) = CQ((d - \frac{n}{2} + 1) mod 4)$$

		$\mathbf{sgn(\widehat{CQ}(d))}$			
d(mod4)	**CQ(d)**	$n \equiv 0(8)$	$n \equiv 2(8)$	$n \equiv 4(8)$	$n \equiv 6(8)$
0	+1	+1	+1	−1	−1
1	+1	−1	+1	+1	−1
2	−1	−1	−1	+1	+1
3	−1	+1	−1	−1	+1

LEMMA 4.2. *Let n be even and let CQ be the complete quadratic function on n variables. Let S be an arbitrary subset of $[n]$ and let $|S| = s$.*

1. *If $(s - \frac{n}{2} + 1) \equiv 0(4))$ then $CQ \cdot \chi_S$ is self-dual bent.*
2. *If $(s - \frac{n}{2} + 1) \equiv 2(4))$ then $CQ \cdot \chi_S$ is anti-self-dual bent.*

	$\mathbf{\|S\|}$ **mod 4**			
$\mathbf{\frac{n}{2}}$**(mod4)**	**0**	**1**	**2**	**3**
0	*X*	*anti*	*X*	*dual*
1	*dual*	*X*	*anti*	*X*
2	*X*	*dual*	*X*	*anti*
3	*anti*	*X*	*dual*	*X*

PROOF. We have already seen that when n is even, CQ is bent and consequently $CQ \cdot \chi_S$ is bent for any S.

Let $g(x) = CQ(x)\chi(x)$. Let x be an arbitrary vertex on the boolean cube and α the corresponding vertex on the dual boolean cube. Given that g is bent, we just need to show that $g(x)$ and $\hat{g}(\alpha)$ have the same(resp. opposite) sign for all x to conclude that g is self dual (resp. anti-self-dual).

Let $|x| = |\alpha| = d$ and $S \cdot x = c$.

$$\begin{aligned}
sign(\hat{g}(\alpha)) &= sign(\widehat{CQ}(\alpha \oplus S)) \\
&= sign(\widehat{CQ}(d + s - 2c) \\
&= sign(CQ((d + s - 2c - \frac{n}{2} + 1) mod 4) \\
&= (-1)^c sign(CQ((d + s - \frac{n}{2} + 1) mod 4) \\
&= (-1)^{\frac{s - \frac{n}{2} + 1}{2}} \cdot (-1)^c \cdot sign(CQ(d)) \\
&= (-1)^{\frac{s - \frac{n}{2} + 1}{2}} \cdot g(x)
\end{aligned}$$

QED. □

5. DEGREE, INFLUENCE, JUNTAS

In this section we only consider functions $f : \{-1,1\}^n \to \mathbb{R}$. For a random variable X (e.g. $f(x)$ where $x \sim \{-1,1\}^n$) we define, $\|X\|_p = \mathbb{E}[|X|^p]^{\frac{1}{p}}$ and analogously $\|f\|_p$.

THEOREM 5.1. **(Exact Junta)** *If $deg(f) = k$ then f is a $k2^{k-1}$-junta. Note that a depth-k decision tree is a 2^{k-1}-junta, so the bound is tight up to the linear factor.*

PROOF. At most $k2^{k-1}$ variables have non-zero influence:

1. For a degree-k function

$$I_i[f] \geq \frac{1}{2^{k-1}} \quad \textit{or} \quad I_i[f] = 0$$

   (a) $I_i[f] = \mathbb{P}[D_i[f] \neq 0]$

   (b) $deg(D_i[f]) = k - 1$

   (c) If $p : \{-1,1\}^n \to \mathbb{R}$ is a degree-k polynomial then

$$p(x) \cong 0 \quad \textit{or} \quad \mathbb{P}[p(x) \neq 0] \geq 2^{-k}$$

      i. Induction on n, k.

      ii. Let $p(x_1, \ldots, x_n, \pm 1) = p_{\pm} = q(x) \pm r(x)$, where $r(x)$ has degree $k-1$.

      iii. If $p_+ \ncong 0$ and $p_- \ncong 0$, done.

      iv. If $p_+ \cong 0$ then p_- has degree $\leq k-1$, done.

2. $I[f] \leq deg(f) = k$.

□

Definition Given $f : \{-1,1\}^n \to \mathbb{R}$, and a *criterion* $g : \mathbb{R} \to \{false, true\}$ we define the **g-slice** of a function as follows:

$$f_g(x) = \begin{cases} f(x) & \text{if } g(f(x)) = true \\ 0 & \text{otherwise} \end{cases}$$

In particular we will frequently use the following shorthand:

$$f_{\geq s}(x) = \begin{cases} f(x) & \text{if } f(x) \geq s \\ 0 & \text{otherwise} \end{cases}$$

$$f_{|\geq s|}(x) = \begin{cases} f(x) & \text{if } |f(x)| \geq s \\ 0 & \text{otherwise} \end{cases}$$

$$|f|_{\geq s}(x) = \begin{cases} |f(x)| & \text{if } |f(x)| \geq s \\ 0 & \text{otherwise} \end{cases}$$

5.1 Notes

1. ***Bonami's Lemma***: degree-k functions are 9^k-reasonable.

   PROOF. $deg(f) = k \implies \|f\|_q \leq \sqrt{q-1}\|f\|_2$ □

2. ***Paley-Zygmund inequality***: For $X \geq 0$ and $0 \leq t \leq 1$, $\mathbb{P}(X \geq t\mu) \geq (1-t)^2 \frac{\mu^2}{\mathbb{E}[X^2]}$

   PROOF.

$$\mu = \mathbb{E}[X] = \sum_{x < t\mu} x\mathbb{P}(x) + \sum_{x \geq t\mu} \sqrt{x^2\mathbb{P}(x)}\sqrt{\mathbb{P}(x)}$$

   Using Cauchy-Schwartz on second term,

$$\mu \leq t\mu + \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{P}[X > t\mu]}$$

   □

3. ***Tail lower bounds***

   (a) ***B-reasonable variables***: If X is B-reasonable,

$$\mathbb{P}[|X| \geq t\|X\|_2] \geq \frac{(1-t^2)^2}{B}$$

   PROOF. Paley-Zygmund applied to X^2. □

   (b) ***Degree-k functions***:

$$\Pr_{x \sim \{-1,1\}^n}[|f(x) - \mu_f| \geq t\sigma_f] \geq \frac{(1-t^2)^2}{9^k}$$

   PROOF. Set $X = \frac{f-\mu}{\sigma}$. Clearly $\|X\|_2 = 1$ and X is 9^k-reasonable. □

4. ***Tail upper bounds***

   (a) ***Hoeffding bound***: Let $X = \sum X_i$, where $X_i \in [a_i, b_i]$. Let $\mu = \mathbb{E}[X]$ and $\lambda^2 = \sum_i (b_i - a_i)^2$

$$\mathbb{P}[|X - \mu| \geq s] \leq 2e^{-2s^2/\lambda^2}$$

   (b) ***Hoeffding for expected value***: Let $\alpha \in \mathbb{R}^n$, $\|\alpha\|_2 = 1$, $x \in \{-1,1\}^n$ and $l(x) = \alpha^T x$.

$$\mathbb{E}[l|_{\geq s}(x)] \leq (2s+2)e^{-s^2/2}$$

   PROOF. Hoeffding bound applied to $\sum X_i$, where each $X_i \in [-\alpha_i, \alpha_i]$

$$\begin{aligned} \mathbb{E}[|l|_{\geq s}(x)] &= s\mathbb{P}[|l(x)| > s] + \int_s^\infty \mathbb{P}[|l(x)| > u]du \\ &\leq 2se^{-s^2/2} + \int_s^\infty 2e^{-s^2/2}du \\ &\leq 2se^{-s^2/2} + \int_s^\infty 2ue^{-s^2/2}du \\ &\leq (2s+2)e^{-s^2/2} \end{aligned}$$

   □

5.2 0-1 bands of a boolean valued function

In this section we show the following properties for boolean valued functions:

1. **Heavy 1-band is unbalanced**: If $1-\epsilon$ of the weight is concentrated in the degree 1 band then most of that weight will be at a single character. (FKN Theorem)

2. **Heavy 0-band $\implies$ light 1-band**: If $W^0[f] \geq 1-\epsilon$, $W^1[f] \leq O(\epsilon^2 \log(1/\epsilon))$.

3. **Balanced 1-band is light**: If none of the singleton fourier coefficients is large (say bounded above by $\epsilon = \frac{1}{100}$) then the degree 1 weight is at most $\frac{2}{\pi} + O(\epsilon)$.

THEOREM 5.2. *(***FKN***) If f is ϵ-concentrated* ***at*** *degree 1 then f is f is $O(\epsilon)$-close to $\pm\chi_i$ for some $i \in [n]$.*

PROOF. Let $l = f^{=1}$ and $\hat{f}_{max} = \max_{1 \leq i \leq n} |\hat{f}_i|$.

1. It is enough to show that $Var[l^2] = O(\epsilon)$.

$$\begin{aligned} l^2 &= (1-\epsilon) + 2\sum_{i<j} \hat{f}_i \hat{f}_j \chi_{i,j} \\ \implies Var[l^2] &= 4\sum_{i<j} \hat{f}_i^2 \hat{f}_j^2 \\ \frac{1}{2} Var[l^2] &= \sum_{i \neq j} \hat{f}_i^2 \hat{f}_j^2 \\ &= (\sum \hat{f}_i^2)^2 - \sum \hat{f}_i^4 \\ \implies \sum \hat{f}_i^4 &= (1-\epsilon)^2 - \frac{1}{2} Var[l^2] \\ \hat{f}_{max}^2 (\sum \hat{f}_i^2) &\geq (1-\epsilon)^2 - \frac{1}{2} Var[l^2] \\ \implies \hat{f}_{max} &\geq (1-\epsilon) - \frac{Var[l^2]}{2(1-\epsilon)} \end{aligned}$$

2. If l^2 is bounded away from 1, then l is bounded away from f: let $\epsilon < 10^{-4}$ and $t = \frac{1}{2}$. Suppose $Var[l^2] > 2601\epsilon$,

$$\begin{aligned} \mathbb{P}[|l^2 - 1 + \epsilon| \geq \frac{51}{2}\sqrt{\epsilon}] &\geq \frac{1}{144} \text{ (using 3b)} \\ \implies \mathbb{P}[|l^2 - 1| \geq \frac{50}{2}\sqrt{\epsilon}] &\geq \frac{1}{144} \\ \implies \mathbb{P}[(f-l)^2 > 146\epsilon] &\geq \frac{1}{144} \\ \implies \mathbb{E}[(f-l)^2] &\geq \frac{146}{144}\epsilon > \epsilon \end{aligned}$$

□

LEMMA 5.3. ***deg*** $\leq$ ***1****: If f is ϵ-concentrated* ***up to*** *degree 1 then f is $O(\epsilon)$-close to a 1-junta (k-junta depends on at most k variables).*

PROOF. Apply FKN to g, defined as:

$$g(x_0, x) = x_0 f(x_0 x) = \hat{f}_\varnothing x_0 + \sum_{|S|=1} \hat{f}_S x^S + \sum_{|S|=2} \hat{f}_S x_0 x^S + \ldots$$

□

THEOREM 5.4. *Let $f : \{-1,1\}^n \to [-1,1]$ and $\alpha = \mathbb{E}[|f|]$. We will assume $\alpha < \frac{1}{2}$, (otherwise consider $1 - f$).*

$$W^1[f] \leq 2\alpha^2 \ln(\frac{1}{\alpha})$$

PROOF. Let $\lambda^2 = W^1[f]$ and $l = \frac{1}{\lambda} f^{=1}$.

$$\begin{aligned} \langle f, l \rangle &= \mathbb{E}(f(x) l_{<s}(x)] + \mathbb{E}(f(x) l_{\geq s}(x)) \\ &\leq \alpha s + 4se^{-s^2/2} \end{aligned}$$

Pick $s = O(ln(1/\alpha))$ □