MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.265/15.070J Fall 2013
Lecture 13 10/23/2013

Concentration Inequalities and Applications

Content.

1 Talagrand's inequality

Let $(\Omega_i, \mathcal{F}_i, \mu_i)$ be probability spaces $(i = 1, ..., n)$. Let $\mu = \mu_1 \bigotimes ... \bigotimes \mu_n$ be product measure on $X = \Omega_1 \times ... \times \Omega_n$. Let $x = (x_1, ..., x_n) \in X$ be a point in this product space.
Hamming distance over X:

$$d(x, y) = |\{i \leq i \leq n : x_i \neq y_i\}| = \sum_{i=1}^{n} \mathbf{1}_{\{x_i \neq y_i\}}$$

α-weighted Hamming distance over X for $a \in \mathbb{R}_+^n$:

$$d_a(x, y) = \sum_{i=1}^{n} a_i \mathbf{1}_{\{x_i \neq y_i\}}$$

Also $|a| = \sqrt{\sum a_i^2}$.
Control-distance from a set: for set $A \subseteq X$, and $x \in X$:

$$\mathcal{D}_A^c(x) = \sup_{|a|=1} d_a(x, A) = \inf\{d_a(x, y) : y \in A\}$$

Theorem 1 (Talagrand)**.** *For every measurable non-emply set A and product-measure μ,*

$$\int \exp(\frac{1}{4}(\mathcal{D}_A^c)^2) d\mu \leq \frac{1}{\mu(A)}$$

In particular,

$$\mu(\{\mathcal{D}_A^c \geq t\}) \leq \frac{1}{\mu(A)} \exp(-\frac{t^2}{4})$$

2 Application of Talagrand's Inequality

2.1 Concentration of Lipschitz functions.

Let $F : X \to \mathbb{R}$ for product space $X = \Omega_1 \times ... \times \Omega_n$ such that for every $x \in X$, there exists $a \equiv a(x) \in \mathbb{R}_+^n$ with $|a| = 1$ so that for each $y \in Y$,

$$F(x) \leq F(y) + d_a(x, y) \tag{1}$$

Why does every 1-Lipschitz function is essentially like (1)?
Consider a 1-Lipschitz function $f : X \to \mathbb{R}$ such that

$$|f(x) - f(y)| \leq \sum_i |x_i - y_i| \text{ (defined on } \Omega_i\text{) for all } x, y \in X.$$

Let $d_i = \max_{x,y\in\Omega} |x_i - y_i|$. We assume d_i is bounded for all i. Then,

$$|f(x) - f(y)| \leq \sum_i |x_i - y_i| \leq \sum_i \mathbf{1}_{\{x_i \neq y_i\}} d_i$$

Therefore,

$$\frac{f(x) - f(y)}{\sqrt{\sum_i d_i^2}} \leq \sum_i \frac{d_i}{\sqrt{\sum d_i^2}} \mathbf{1}_{\{x_i \neq y_i\}} = d_a(x, y) \text{ with } a_i = \frac{d_i}{\sqrt{\sum_i d_i^2}}$$

Thus $F(x) = \frac{f(x)}{||d||_2}$ where $||d||_2 = \sqrt{\sum_i d_i^2}$.

Let $A = \{F \leq m\}$. By definition of $\mathcal{D}_A^c(x)$,

$$\mathcal{D}_A^c(x) = \sup_{a:|a|=1} d_a(x, A) \geq d_a(x, y)$$

for a given a such that $|a| = 1$ and $y \in A$. Now for any $y \in A$, by definition $F(y) \leq m$. Then,

$$F(x) \leq F(y) + d_a(x, y) \leq m + \mathcal{D}_A^c(x)$$

which implies $\{F \geq m + r\} \subseteq \{\mathcal{D}_A^c(x) \geq r\}$. By Talagrand's inequality, for any $r \geq 0$,

$$\mathbb{P}(\{f \geq m + r\}) \leq \mathbb{P}(\{\mathcal{D}_A^c \geq r\}) \leq \frac{1}{\mathbb{P}(A)} \exp(-\frac{r^2}{4})$$

That is,

$$\mathbb{P}(\{F \leq m\})\mathbb{P}(\{F \geq m + r\}) \leq \exp(-\frac{r^2}{4}) \tag{2}$$

The median of F, m_F is precisely such that

$$\mathbb{P}(F \leq m_F) \geq \frac{1}{2},\ \mathbb{P}(F \geq m_F) \geq \frac{1}{2}$$

Choose $m = m_F, m = m_F - r$ in (2) to obtain:

$$\mathbb{P}(F \geq m_F + r) \leq 2\exp(-\frac{r^2}{4}),\ \mathbb{P}(F \leq m_F - r) \leq 2\exp(-\frac{r^2}{4}) \quad (3)$$

Thus,

$$\mathbb{P}(|F - m_F| \geq r) \leq 4\exp(-\frac{r^2}{4})$$

2.2 Further Application for Linear Functions

Consider the independent random variables $Y_1, ..., Y_n$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let the constants (u_i, v_i), $1 \leq i \leq n$ such that

$$u_i \leq Y_i \leq v_i$$

Set $Z = \sup_{t\in T} < t, Y > \equiv \sum_{i=1}^n t_i Y_i$ where T is some finite, countable or compact set of vectors in $\mathbb{R}_+$. We would be interested in situations where

$$\sigma^2 = \sup_{t\in T} \sum_i t_i^2 (v_i - u_i)^2 \leq \infty$$

We wish to apply (3) to this setting by choosing

$$F(x) = \sup_{t\in T} < t, x >$$

where $x \in X$ and $X = \prod_{i=1}^n [u_i, v_i]$. Given that T is compact, $F(x) =< t^*(x), x >$ for some $t = t^*(x) \in T$, given x.

$$\begin{aligned} F(x) = \sum_{i=1}^n t_i x_i &\leq \sum_i t_i y_i + \sum_i |t_i||y_i - x_i| \\ &\leq \sum_i t_i y_i + \sum_i |t_i|(v_i - u_i)\mathbf{1}_{(y_i \neq x_i)} \text{ (let } d_i = |t_i|(v_i - u_i)) . \\ &\leq \sup_{\tilde{t}\in T} < \tilde{t}, y > + (\sum_i \frac{d_i}{||d||_2}\mathbf{1}(y_i \neq x_i))||d||_2 \\ &= F(y) + d_a(x, y)||d||_2 \text{ (where let } \sigma = ||d||_2 = \sqrt{\sup_{t\in T} \sum t_i^2 (v_i - u_i)^2}) \\ &= F(y) + \sigma d_a(x, y) \quad (4) \end{aligned}$$

By selection of $f \equiv \frac{1}{\sigma}F$, (3) can be applied to f:

$$\mathbb{P}(|f - m_f| \geq r) \leq 4\exp(-\frac{r^2}{4})$$

Let $r = \frac{\gamma}{\sigma}$, then $\mathbb{P}(|\sigma f - \sigma m_f| \geq \gamma) \leq 4\exp(-\frac{\gamma^2}{4\sigma^2})$. That is,

$$\mathbb{P}(|F - m_F| \geq \gamma) \leq 4\exp(-\frac{\gamma^2}{4\sigma^2})$$

Now,

$$\begin{aligned}
\mathbb{E}[F] &= \int_0^\infty \mathbb{P}(F \geq s)ds \text{ (assume } t \equiv 0 \in T) \\
&\leq \int_0^{m_F} 1ds + \int_0^\infty \mathbb{P}(F \geq m_F + \gamma)d\gamma \\
&\leq m_F + \int_0^\infty 2\exp(-\frac{\gamma}{4\sigma^2})d\gamma \\
&\leq m_F + \int_0^\infty 2\exp(-\frac{\gamma^2}{4\sigma^2})d\gamma \\
&= m_F + 2\sqrt{8\pi\sigma^2}\int_0^\infty \frac{1}{\sqrt{2\pi 4\sigma^2}}\exp(-\frac{\gamma^2}{4\sigma^2})d\gamma \\
&= m_F + 2\sqrt{2\pi}\sigma
\end{aligned}$$

Thus,

$$|\mathbb{E}[F] - m_F| \leq 2\sqrt{2\pi}\sigma$$

2.3 More Intricate Application

Longest increasing subsequence:
Let $X_1, ..., X_n$ be points in $[0, 1]$ chosen independently as a product measure. Let $L_n(X_1, ..., X_n)$ be the length of longest increasing subsequence. (Note that $L_n(\cdot)$ is not obviously Lipschitz). Talagrand's inequality implies its concentration.

Lemma 1. *Let m_n be median of L_n. Then for any $r > 0$, we have*

$$\mathbb{P}(L_n \geq m_n + r) \leq 2\exp(-\frac{r^2}{4(m_n + r)})$$

$$\mathbb{P}(L_n \leq m_n - r) \leq 2\exp(-\frac{r^2}{4m_n})$$

Proof. Let us start by establishing first inequality. Select $A = \{L_n \leq m_n\}$. Clearly, by definition $\mathbb{P}(A) \geq \frac{1}{2}$. For a x such that $L_n(x) > m_n$, (i.e. $x \in A$), consider any $y \in A$. Now, let set $I \subseteq [n]$ be indices that give rise to longest increasing subsequence in x: i.e. say $I = \{i_1, ..., i_p\}$ then $x_{i_1} < x_{i_2} < ... < x_{i_p}$ and p is the maximum length of any such increasing subsequence of x. Let $J = \{i \in I : x_i \neq y_i\}$ for given y. Since $I \backslash J$ is an index set that corresponds to a increasing subsequence of y (since for $i \in I \backslash J$; $x_i = y_i$ and I is index set of increasing subsequence of I); we have that (using fact that $L_n(y) \leq m_n$ as $y \in A$)

$$|I \backslash J| \leq m_n$$

That is,

$$\begin{aligned} L_n(x) = |I| &\leq |I \backslash J| + |J| \\ &\leq L_n(y) + \sum_{i \in I} \mathbf{1}(x_i \neq y_i) \\ &\leq L_n(y) + \sqrt{L_n(x)}[\sum_{i=1}^{n} \frac{1}{\sqrt{L_n(x)}} \mathbf{1}(i \in I)\mathbf{1}(x_i \neq y_i)] \end{aligned}$$

Define

$$a_i(x) = \begin{cases} \frac{1}{\sqrt{L_n(x)}}, & \text{if } i \in I \\ 0, & \text{o.w.} \end{cases}$$

Then $|a| = 1$ since $|I| = L_n(x)$ by definition, and hence,

$$L_n(x) \leq L_n(y) + \sqrt{L_n(x)} d_a(x, y) \leq m_n + \sqrt{L_n(x)} \mathcal{D}_A^c(x)$$

Equivalently,

$$\mathcal{D}_A^c(x) \geq \frac{L_n(x) - m_n}{\sqrt{L_n(x)}}$$

For x such that $L_n(x) \geq m_n + r$, the RHS of ahove is minimal when $L_n(x) = m_n + r$. Therefore, we have

$$\mathcal{D}_A^c(x) \geq \frac{L_n(x) - m_n}{\sqrt{L_n(x)}}$$

For x such that $L_n(x) \geq m_n + r$, the RHS of above is minimal when $L_n(x) = m_n + r$. Therefore, we have

$$\mathcal{D}_A^c(x) \geq \frac{r}{\sqrt{m_n + r}}$$

That is

$$L_n(x) \geq m_n + r \Rightarrow \mathcal{D}_A^c(x) \geq \frac{r}{\sqrt{m_n + r}} \text{ for } A = \{L_n \leq m_n\}$$

Putting these together, we have

$$\mathbb{P}(L_n \geq m_n + r) \leq \mathbb{P}(\mathcal{D}_A^c \geq \frac{r}{\sqrt{m_n + r}}) \leq \frac{1}{2P(A)} \exp(-\frac{r^2}{4(m_n + r)})$$

But $\mathbb{P}(A) = \mathbb{P}(L_n \leq m_n) \geq \frac{1}{2}$, we have that

$$\mathbb{P}(L_n \geq m_n + r) \leq 2 \exp(-\frac{r^2}{4(m_n + r)})$$

To establish lower bound, replace argument of the above with x such that $L_n(x) \geq s + u$, $A = \{L_n \leq s\}$. Then we obtain,

$$\mathcal{D}_A^c(x) \geq \frac{u}{\sqrt{s + u}}$$

Select $s = m_n - r$, $u = r$. Then whenever x is such that $L_n(x) \geq s + u = m_n$ and for $A = \{L_n \leq s\} = \{L_n \leq m_n - r\}$.

$$\mathcal{D}_A^c(x) \geq \frac{r}{\sqrt{m_n}}$$

Thus,

$$\mathbb{P}(L_n \geq m_n) \leq \mathbb{P}(\mathcal{D}_A^c \geq \frac{r}{m_n}) \leq \frac{1}{\mathbb{P}(L_n \leq m_n - r)} \exp(-\frac{r^2}{4m_n})$$

which implies

$$\mathbb{P}(L_n \leq m_n - r) \leq 2 \exp(-\frac{r^2}{4m_n})$$

This completes the proof. □

3 Proof of Talagrand's Inequality

Preparation. Given set $A, x \in X$: $\mathcal{D}_A^c(x) = \sup_{a \in \mathcal{R}_+^n}(d_a(x, A) = \inf_{y \in A} d_a(x, y))$. Let

$$U_A(x) = \{s \in \{0,1\}^n : \exists y \in A \text{ with } s \triangleq \mathbf{1}(x \neq y)\} = \{\mathbf{1}(x \neq y) : y \in A\}$$

and let

$$V_A(x) = \text{Convex-hull}(U_A(x)) = \{\sum_{s\in U_A(x)} \alpha_s S : \sum \alpha_s = 1, \alpha_s \geq 0 \text{ for all } s \in U_A(x)\}$$

Thus,

$$x \in A \Leftrightarrow \mathbf{1}(x \neq x) = 0 \in U_A(x) \Leftrightarrow 0 \in V_A(x)$$

It can therefore be checked that

Lemma 2.

$$\mathcal{D}_A^c(x) = d(0, V_A(x)) \equiv \inf_{y\in V_A(x)} |y|$$

Proof. (i) $\mathcal{D}_A^c(x) \leq \inf_{y\in V_A(x)} |y|$: since $\inf_{y\in V_A(x)}(y)$ is achieved, let Z be such that $|Z| = \inf_{y\in V_A(x)} |y|$. Now for any $a \in \mathbb{R}_+^n$, $|a| = 1$:

$$\inf_{y\in V_A(x)} a\cdot y \leq a\cdot z \leq |a||z| = |z|$$

Since $\inf_{y\in V_A(x)} a\cdot y$ is linear programming, the minimum is achieved at an extreme point. That is, there exists $s \in U_A(x)$ such that

$$\inf_{y\in V_A(x)} a\cdot y = \inf_{s\in U_A(x)} a\cdot s = \inf_{y\in A} d_a(x,y) \text{ for some } y \in A.$$

Since this is true for all a, it follows that,

$$\sup_{|a|=1, a\in\mathbb{R}_+^n} \inf_{y\in A} d_a(x,y) \leq |z| \equiv \inf_{y\in V_A(x)} |y|$$

(ii) $\mathcal{D}_A^c(x) \geq \inf_{y\in V_A(x)} |y|$: Let z be the one achieving minimum in $V_A(x)$. Then due to convexity of the objective (equivalently $|y|^2 = \sum y_i^2 = f(y)$) and of the domain, we have for any $y \in V_A(x)$, $\triangledown f(z)(y-z) \geq 0$ for any $y \in V_A(x)$. $\triangledown f(z) = \triangledown(z\cdot z) = 2z$. Therefore the condition implies

$$(y-z)z \geq 0 \Leftrightarrow y\cdot z \geq z\cdot z = |z|^2 \Rightarrow y\cdot \frac{z}{|z|} \geq |z|$$

Thus, for $a = \frac{z}{|z|} \in \mathbb{R}_+^n$, $|a| = 1$, we have that

$$\inf_{y\in V_A(x)} a\cdot y \geq |z|$$

But for any given a, $\inf_{y\in V_A(x)} a\cdot y = \inf_{s\in U_A(x)} a\cdot s = d_a(x,A)$ as explained before. That is, $\sup_{a:|a|=1} d_a(x,A) \geq |z| = \inf_{y\in V_A(x)} |y|$. This completes the proof. □

Now we are ready to establish the inequality of Talagrand. The proof is via induction. Consider $n = 1$, given set A. Now,

$$\mathcal{D}_A^c(x) = \sup_{a \in \mathbb{R}_+^n, |a|=1} \inf_{y \in A} d_a(x, y) = \inf_{y \in A} \mathbf{1}(x \neq y) = \begin{cases} 0, & \text{for } x \in A \\ 1, & \text{for } x \notin A \end{cases}$$

Then,

$$\begin{aligned} \int \exp(D^2/4) dP &= \int_A \exp(0) dP + \int_{A^c} \exp(1/4) dP \\ &= P(A) + e^{1/4}(1 - P(A)) \\ &= e^{1/4} - (e^{1/4} - 1)P(A) \leq \frac{1}{P(A)} \end{aligned} \tag{5}$$

Let $f(x) = e^{1/4} - (e^{1/4} - 1)x$ and $g(x) = \frac{1}{x}$. Because $f(x)$ is a decreasing function of x, $g(x)$ is a decreasing convex function. Thus, the result if established for $n = 1$.

Induction hypothesis. Let it hold for some n. We shall assume for ease of the proof that $\Omega_1 = \Omega_2 = ... = \Omega_n = ... = \Omega$. L

Let $A \subset \Omega^{n+1}$. Let B be its projection on Ω^n. Let $A(\omega)$, $\omega \in \Omega$ be section of A along ω: if $x \in \Omega^n$, $\omega \in \Omega$ then $z = (x, \omega) \in \Omega^{n+1}$. We observe the following:

if $s \in U_{A(\omega)}(x)$, then $(s, 0) \in U_A(z)$. Because, for some $y \in \Omega^n$ such that $(y, \omega) \in A$, $s = \mathbf{1}(x \neq y)$. Therefore, $(s, 0) = (\mathbf{1}(x \neq y), \mathbf{1}(\omega \neq \omega)) = \mathbf{1}(z \neq (y, \omega))$ where $(y, \omega) \in A$. Further, if $t \in U_B(x)$, then $(t, 1) \in U_A(z)$. This is because of the following: $B = \{\tilde{x} \in \Omega^n : (\tilde{x}, \tilde{\omega}) \in A \text{ for some } \tilde{\omega} \in \Omega\}$. Now if $t \in U_B(x)$, then $\exists\ y \in B$ such that $t = \mathbf{1}(x \neq y)$. Now $(t, 1) = (\mathbf{1}(x \neq y), \mathbf{1}(\tilde{\omega} \neq \omega)) = \mathbf{1}(z \neq (y, \tilde{\omega}))$ as long as there exists $\tilde{\omega}$ so that $(y, \tilde{\omega}) \in A$ and $\tilde{\omega} \neq \omega$.

Given this, it follows that if $\xi \in V_{A(\omega)}(x)$, $\zeta \in V_B(x)$, and $\theta \in [0, 1]$, then $((\theta\xi + (1 - \theta)\zeta), 1 - \theta) \in V_A(z)$. Recall that

$$\begin{aligned} \mathcal{D}_A^c(z)^2 = \inf_{y \in V_A(z)} |y|^2 &\leq (1 - \theta)^2 + |\theta\xi + (1 - \theta)\zeta|^2 \\ &\leq (1 - \theta)^2 + \theta|\xi|^2 + (1 - \theta)|\zeta|^2 \end{aligned} \tag{6}$$

Therefore,

$$\begin{aligned} \mathcal{D}_A^c(z)^2 &\leq (1 - \theta)^2 + \theta \inf_{\xi \in V_{A(\omega)}(x)} |\xi|^2 + (1 - \theta) \inf_{\zeta \in V_B(x)} |\zeta|^2 \\ &= (1 - \theta)^2 + \theta\mathcal{D}_{A(\omega)}^c(x)^2 + (1 - \theta)\mathcal{D}_B^c(x)^2 \end{aligned}$$

By Hölder's inequality, and the induction hypothesis, for $\forall \omega \in \Omega$,

$$\begin{aligned}
&\int_{\Omega^n} e^{\mathcal{D}_A^c(x,\omega)^2/4} dP(x) \\
&\leq \int_{\Omega^n} \exp\left(\frac{(1-\theta)^2 + \theta\mathcal{D}^c_{A(\omega)}(x)^2 + (1-\theta)\mathcal{D}^c_B(x)}{4}\right) dP(x) \\
&\leq \exp(\frac{(1-\theta)^2}{4}) \int_{\Omega^n} \underbrace{\exp(\frac{\theta\mathcal{D}^c_{A(\omega)}(x)^2}{4})}_{X} \underbrace{\exp(\frac{(1-\theta)\mathcal{D}^c_B(x)^2}{4})}_{Y} dP(x) \\
&= \exp(\frac{(1-\theta)^2}{4})\mathbb{E}[X \cdot Y] \\
&\leq \exp(\frac{(1-\theta)^2}{4})\mathbb{E}[X^p]^{1/p}\mathbb{E}[Y^q]^{1/q}, \text{ (for } p = \frac{1}{\theta}, q = \frac{1}{1-\theta} \text{: } \theta \in [0,1]) \\
&= \exp(\frac{(1-\theta)^2}{4}) \left(\int_{\Omega^n} \exp(\mathcal{D}^c_{A(\omega)}(x)^2/4) dP(x)\right)^\theta \left(\int_{\Omega^n} \exp(\mathcal{D}^c_B(x)^2/4) dP(x)\right)^{1-\theta} \\
&\leq \exp(\frac{(1-\theta)^2}{4})(\frac{1}{P(A(\omega))})^\theta(\frac{1}{P(B)})^{1-\theta} \text{ by induction hypothesis.} \\
&= \exp\left(\frac{(1-\theta)^2}{4}\right) \frac{1}{P(B)} \left(\frac{P(A(\omega))}{P(B)}\right)^{-\theta} \qquad (7)
\end{aligned}$$

(7) is true for any $\theta \in [0,1]$, so for tightest upper bound, we shall optimize.
Claim: for any $u \in [0,1]$, $\inf_{\theta \in [0,1]} \exp(\frac{(1-\theta)^2}{4})u^{-\theta} \leq 2 - u$.
Therefore, (7) reduces to

$$\leq \frac{1}{P(B)}(2 - \frac{P(A(\omega))}{P(B)})$$

Therefore,

$$\begin{aligned}
&\int_{\Omega^{n+1}} \exp(\frac{\mathcal{D}^c_A(x,\omega)^2}{4}) dP(x) d\mu(\omega) \\
&\leq \frac{1}{\mathbb{P}(B)} \int_\Omega (2 - \frac{\mathbb{P}(A(\omega))}{\mathbb{P}(B)}) d\mu(\omega) \\
&\leq \frac{1}{\mathbb{P}(B)}(2 - \frac{(P \bigotimes \mu)(A)}{\mathbb{P}(B)}) \\
&\leq \frac{1}{(\mathbb{P} \bigotimes \mu)(A)}, \text{(since } u(2-u) \leq 1 \text{ for all } u \in \mathbb{R}) \qquad (8)
\end{aligned}$$

This completes the proof of Talagrand's inequality.
Claim: $f(u) = u(2-u) \Rightarrow f^{'}(u) = 2 - 2u \Rightarrow u^* = 1 \Rightarrow \max_u f(u) =$

$f(1) = 1$.

Proof. To establish: $\inf_{\theta \in [0,1]} \exp(\frac{(1-\theta)^2}{4}) u^{-\theta} \leq 2 - u$:
if $u \geq e^{-1/2}$: $\theta = 1 + 2\log u \Rightarrow \frac{1-\theta}{2} = -\log u \Rightarrow \frac{(1-\theta)^2}{4} = \log^2(u)$ and $u^{-\theta} = e^{-\theta \log u} = e^{-\log u} e^{-2\log^2 u}$. Thus,

$$\exp(\frac{(1-\theta)^2}{u}) u^{-\theta} = \exp(\log^2 u - 2\log^2 u - \log u) = \exp(-\log u - \log^2 u)$$

We have that

$$1 \geq u \geq e^{-1/2} \Rightarrow 0 \geq \log u \geq -\frac{1}{2} \Rightarrow 0 \leq -\log u \leq \frac{1}{2},\ 0 \leq \log^2 u \leq \frac{1}{4}$$

and

$$f(x) = -x - x^2 :\ x \in [-1/2, 0];\ f^{'}(x) = -1 - 2x \leq 0 \text{ for } x \in [-1/2, 0]$$

Thus,

$$-\log u - \log^2 u \leq \frac{1}{2} - \frac{1}{4} \leq \frac{1}{4} \Rightarrow \exp(-\log u - \log^2 u) \leq \frac{1}{4}$$

and for $u \geq e^{-\frac{1}{2}}$ which implies that $2 - u \geq \exp(-\log u - \log^2 u)$. □

15.070J / 6.265J Advanced Stochastic Processes
Fall 2013