BEST CONSTANTS IN CHEBYSHEV INEQUALITIES WITH VARIOUS APPLICATIONS

by

Anirban DasGupta

Technical Report #98-20

Department of Statistics
Purdue University
West Lafayette, IN USA

November 1998
Revised February 2000

*Research supported by a National Security Agency Grant

BEST CONSTANTS IN CHEBYSHEV INEQUALITIES
WITH VARIOUS APPLICATIONS

by

Anirban DasGupta
Purdue University

Abstract

In this article we describe some ways to significantly improve the Markov-Gauss-Camp-Meidell inequalities and provide specific applications. We also describe how the improved bounds are extendable to the multivariate case. Applications include explicit finite sample construction of confidence intervals for a population mean, upper bounds on a tail probability $P(X > k)$ by using the density at k, approximation of P-values, simple bounds on the Riemann Zeta function, on the series $\sum_{\text{prime}} \frac{e^{-\lambda p}}{p}$, improvement of Minkowski moment inequalities, and construction of simple bounds on the tail probabilities of asymptotically Poisson random variables. We also describe how a game theoretic argument shows that our improved bounds always approximate tail probabilities to any specified degree of accuracy.

1. Introduction.

The Markov inequality, which states that for any random variable X and positive numbers $k, r, P(|X| > k) \leq \frac{E|X|^r}{k^r}$, is generally regarded as an inefficient bound to be used only when a crude bound suffices. Gauss showed that if X is unimodal with a mode at zero, then the bound can be improved to $P(|X| > k) \leq \frac{4}{9}\frac{E(X^2)}{k^2}$ when $r = 2$. The Camp-Meidell inequality generalizes Gauss's result as $P(|X| > k) \leq (\frac{r}{r+1})^r \frac{E|X|^r}{k^r}$ for all $r > 0$. Sellke and Sellke (1997) provide generalizations to these bounds by using moments of more general even functions $g(|X|)$. The Markov and the Camp-Meidell bounds are sharp in the sense that without further restrictions, the bounds are known to be attained, although at uninteresting distributions; see Dharmadhikari and Joag-dev (1988).

And so an interesting question emerges: for specific important distributions or specific important families of distributions, can we significantly improve these probability bounds? Our intention is to show that this is indeed the case and that, interestingly, a bit of game theory shows that by a judicious choice of the number r, one can approximate the probability $P(|X| > k)$ to any degree of accuracy. For instance, we shall show that if $X \sim N(\mu, \sigma^2)$, then $P(|X - \mu| > k\sigma) < \frac{1}{3k^2}$ for any $k > 0$; indeed, as we shall see, this improvement over Gauss's bound holds for a much larger family of distributions, and is suitably extendable to the multivariate case. Our results also permit very significant sharpening of the Jensen type bound $E|X - \mu| \leq (E|X - \mu|^r)^{\frac{1}{r}}$ for $r > 1$, and lead to simple and useful bounds of the type $P(X > k) \leq C(k)f(k)$ where $C(k)$ is an explicit constant and $f(k)$ the density at k.

For discrete unimodal distributions, the Gauss-Camp-Meidell bounds are not applicable. For this reason, we shall also provide appropriate analogous results for an important discrete case, namely the Poisson distribution. Here there will be simple bounds using the easily available Stirling and Bell numbers.

The results are illustrated with some other specific applications such as the number of fixed points of a random permutation of $1, 2, \ldots, n$, approximation of multivariate probabilities, and construction of explicit finite sample confidence intervals for a population mean μ. For example, we have the general result that in the entire normal scale mixture family, $\overline{X} \pm 1.82\frac{\sigma}{\sqrt{n}}$ is always a 90% confidence interval for the population mean μ, for every n. We also give some analytic bounds on certain number-theoretic functions that follow from our main theorem.

2. 2.1. The Markov Bound.

The precise question asked in this section is the following. Let a real valued random variable X be distributed according to some specified CDF F. For given $r \geq 0$, what is the exact best possible value of a constant $C(r) = C(F, r)$ such that $P(|X| > k) \leq C(r) \frac{E_F|X|^r}{k^r}$ for any $k > 0$, and how to find analytic good bounds on this best possible constant. In addition, since we can choose r to be any nonnegative number, can we always well approximate $P(|X| > k)$ by a judicious choice of r?

Notation. The best possible constant will be denoted by

$$C^*(r) = C^*(F, r).$$

Theorem 1.

a) $C^*(r) = \dfrac{\sup\limits_{x>0}\{x^r P_F(|X|>x)\}}{E_F|X|^r}$ (2.1)

b) If $|X|$ has a density f, then $C^*(r) \leq \dfrac{\sup\limits_{x>0}\{x^{r+1} f(x)\}}{r E_F|X|^r} \overset{\text{def}}{=} \overline{C}(r)$ (2.2)

c) If F is symmetric and absolutely continuous with characteristic function $\psi(t)$, then $C^*(r)$ is also equal to $\frac{2}{\pi E_F|X|^r} \sup_{x>0}\{x^r \int\limits_0^\infty \frac{\sin tx}{t}(1 - \psi(t))dt\}$ (2.3)

Discussion. Part a) of Theorem 1 gives a general exact expression for $C^*(r)$, but the criticism can be raised that it involves the very quantities one is trying to approximate in the first place. This is true; however, if $C^*(r)$ is once computed, then the resulting improved probability bounds may be useful to future researchers. Parts b) and c) provide alternative ways to compute or bound $C^*(r)$. If a formula for the density is available, then part b) is very useful as $\overline{C}(r)$ will often be closed form; we shall later see that this bound on $C^*(r)$ is a pretty good bound considering its simplicity. In some problems, however, a formula for a density will not be available; difficult convolutions (e.g., convolutions of t densities) are instances of such cases. In such cases, part c) can become useful due to the availability of a formula for a characteristic function.

Proof of Theorem 1: Part a) is essentially a restatement of the definition of $C^*(r)$ and does not need a proof. To prove part b), note that the supremum is attained at a point $x_* = x_*(F, r)$ such that $P(|X| > x_*) = \frac{x_* f(x_*)}{r}$ and hence $\sup_{x>0}\{x^r P(|X| > x)\} = \frac{x_*^{r+1} f(x_*)}{r} \leq \frac{\sup_{x>0}\{x^{r+1} f(x)\}}{r}$ and so this proves part b).

For part c), note that if F is symmetric and absolutely continuous, then $F(x) = \frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin tx}{t} \psi(t) dt$. Combine this with the familiar fact that $\int_{-\infty}^{\infty} \frac{\sin u}{u} du = \pi$ to get

$$
\begin{aligned}
& P(|X| > x) \\
&= 2(1 - F(x)) = 1 - \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin tx}{t} \psi(t) dt \\
&\qquad = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin tx}{t} dt - \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin tx}{t} \psi(t) dt \\
&\qquad = \frac{2}{\pi} \int_{0}^{\infty} \frac{\sin tx}{t} (1 - \psi(t)) dt,
\end{aligned}
$$

and this proves part c).

Example 1. Comparison to Camp-Meidell in the Normal Case. Recall that the Camp-Meidell inequality says that $(P|X| > k) \leq (\frac{r}{r+1})^r \frac{E|X|^r}{k^r}$ if X is unimodal with a mode at 0. Let $X \sim N(0, \sigma^2)$ as an example. The value of $C^*(r)$ is the same for all σ and so let us take $\sigma = 1$. The following table compares $C^*(r)$ with $\overline{C}(r)$ and $(\frac{r}{r+1})^r$. Note that $\overline{C}(r)$ has the closed form expression

$$
\overline{C}(r) = \frac{\frac{2}{r} (\frac{r+1}{2})^{\frac{r+1}{2}} e^{-\frac{r+1}{2}}}{\Gamma(\frac{r+1}{2})} \tag{2.4}
$$

TABLE 1

r	$C^*(r)$	$\overline{C}(r)$	$(\frac{r}{r+1})^r$
1	.426	.736	.5
2	.331($< \frac{1}{3}$)	.463	.444
3	.283	.361	.422
4	.251	.305	.410
6	.212	.243	.397
8	.187	.208	.390

For $r \geq 3, \overline{C}(r)$ is smaller than the Camp-Meidell constant and for $r \geq 4$ it is quite close to the exact best constant $C^*(r)$. Note that from TABLE 1, if $X \sim N(\mu, \sigma^2)$, then $P(|X - \mu| > k\sigma) < \frac{1}{3k^2}$ for all $k > 0$. Since this bound will work for all scale mixtures as well, one has the interesting consequence:

Proposition 1.

Let $X_1, X_2, \ldots, X_n$ be n iid observations from any distribution in the normal scale mixture family with $E(X_1) = \mu, Var(X_1) = \sigma^2$. Then

$$\begin{cases} \overline{X} \pm 2.58\frac{\sigma}{\sqrt{n}} & \text{is always a 95\% confidence interval for } \mu \\ \overline{X} \pm 1.82\frac{\sigma}{\sqrt{n}} & \text{is always a 90\% confidence interval for } \mu \\ \overline{X} \pm .81\frac{\sigma}{\sqrt{n}} & \text{is always a 50\% confidence interval for } \mu \end{cases}$$

We simulated the coverage probability of the interval $\overline{x} \pm 2.58\frac{s}{\sqrt{n}}$ for five normal scale mixture densities when $n = 10$. s was used instead of σ as σ would usually be unknown in an application. The simulated coverage probabilities (using a simulation size of 15,000) are as follows:

Density	**Coverage**
Double Exp.	.9752
Logistic	.9721
$t(3)$	.9732
$t(5)$	.9712
$\cdot 9N(0,1) + \cdot 1N(0,9)$	.9774

The coverage seems very stable (between .97 and .98 in each case), but conservative. What we are seeing is that t-intervals work well for many symmetric densities. But, of course, a result such as Proposition 1 cannot be proved for the t-interval.

Let us check the efficacy of the bound $P(|X| > k) \leq C^*(r)\frac{E|X|^r}{k^r}$ in but one case to motivate our next Theorem. Take X to be $N(0,1)$ and $k = 2$. Then $P(|X| > k) = .04550$. The bound $C^*(r)\frac{E|X|^r}{k^r}$ is listed below for several choices of r.

r	1	2	3	4	6
Bound	.16997	.08286	.05642	.04715	.04962

These numbers suggest that the bound $C^*(r)\frac{E|X|^r}{k^r}$ could be a very good approximation to $P(|X| > k)$ if r is chosen properly. The next Theorem says that is indeed the case.

Theorem 2.

Suppose for some $r_0 > 0$, $\lim_{x\to\infty} x^{r_0} P(|X| > x) = 0$. Let $B(k, r) = C^*(r)\frac{E|X|^r}{k^r}$. Then $\inf_{r>0} B(k, r) = P(|X| > k)$ for any $k > 0$.

Proof: Consider a game with nature's action space $= \Theta = [0, \infty]$, statistician's action space $= \mathcal{A} = (0, \infty)$, and statistician's payoff function $L(\theta, a) = (\frac{\theta}{k})^a P(|X| > \theta)$. This payoff function is upper semicontinuous in θ for fixed 'a', $\inf_a \sup_\theta L(\theta, a)$ is finite and the game has a value under the stated assumption

$$\lim_{x\to\infty} |x|^{r_0} P(|X| > x) = 0.$$

$$\begin{aligned}
\text{Thus, } & \inf_{r>0} B(k, r) \\
&= \inf_{r>0} \frac{\sup_{x\geq 0} x^r P(|X| > x)}{k^r} \\
&= \inf_{r>0} \sup_{x\geq 0} L(x, r) \\
&= \sup_{x\geq 0} \inf_{r>0} L(x, r) \\
&= \max\{\sup_{x<k} \inf_{r>0} L(x, r), \sup_{x\geq k} \inf_{r>0} L(x, r)\} \\
&= \max\{0, \sup_{x\geq k} P(|X| > x)\} \\
&= P(|X| > k),
\end{aligned}$$

as stated.

Remark.

Theorem 2 says that $P(|X| > k)$ is always reproduced exactly by the bound $B(k, r)$ if r is chosen properly. The question arises if a guide exists for choosing r. Part b) of Theorem 1 gives an indication. A guide to choosing r is to find the value of r that minimizes $\frac{1}{rk^r} \sup_{x>0} x^{r+1} f(x)$. Here is a quick test case.

Example 2. Choice of the optimal r.

Let X have the standard double exponential density $f(x) = \frac{1}{2}e^{-|x|}$. Then the bound $B(k, r)$ equals $\frac{r^r e^{-r}}{k^r}$, which is minimized when $r = k$. On the other hand, $\frac{1}{rk^r} \sup_{x>0} x^{r+1} f(x)$

$= \frac{1}{2e}\frac{r+1}{r}\frac{(r+1)^r e^{-r}}{k^r}$, which is minimized when $\log(1+r) - \frac{1}{r} = \log k$. This equals $1.24, 2.17, 3.13, 4.10, 5.09$ for $k = 1, 2, 3, 4, 5$ respectively. So in this case, the approximation to the optimal r is quite good for moderate or large k.

2.2. Applications in Mathematics.

The inequality of Theorem 1 which states that if $X \sim F$, then $P_F(|X| > k) \leq C^*(r)\frac{E_F|X|^r}{k^r}$ can lead to useful and interesting bounds for mathematical functions. The method is to select a specific F, derive explicitly the above inequality for that F, and then sum the inequalities for the values of k in a given set, say, $1 \leq k \leq n$. One can then further select optimal or special values of r to get a more refined inequality.

2.2.1. Bounds on the Riemann Zeta Function.

Let X have the Geometric distribution with parameter p, i.e., $P(X = x) = pq^x, x \geq 0$, where $q = 1 - p$. Then, from Theorem 1 it follows that

$$q^{k+1} = P(X > k) \leq \frac{(\frac{-r}{\log q})^r q^{1+[\frac{-r}{\log q}]}}{k^r}, \tag{2.5}$$

where $[\cdot]$ denotes the integer part.

From (2.5), $\forall r > 0, 0 < q < 1$,

$$\frac{1}{k^r} \geq \frac{q^{k+1}}{(\frac{-r}{\log q})^r q^{1+[\frac{-r}{\log q}]}},$$

and in particular, if q is written as $q = e^{-\alpha}$ for some positive number α, then

$$\frac{1}{k^r} \geq \frac{e^{-(k+1)\alpha}}{(\frac{r}{\alpha})^r e^{-\alpha(1+[\frac{r}{\alpha}])}} \tag{2.6}$$

$$\Rightarrow \sum_{k=1}^{\infty}\frac{1}{k^r} \geq \frac{\sum_{k=1}^{\infty} e^{-(k+1)\alpha}}{(\frac{r}{\alpha})^r e^{-\alpha(1+[\frac{r}{\alpha}])}} = \frac{\alpha^r e^{\alpha[\frac{r}{\alpha}]}}{(e^\alpha - 1)r^r}, \tag{2.7}$$

on summing the geometric series of the numerator.

In (2.7), if we now let $\alpha = r(> 1)$, then for the Riemann Zeta function $\zeta(\alpha)$, we get the simple bound:

$$\zeta(\alpha) = \sum_{k=1}^{\infty} \frac{1}{k^{\alpha}} \geq \frac{e^{\alpha}}{e^{\alpha} - 1}, \tag{2.8}$$

for any α.

The following table illustrates the sharpness of the bound.

TABLE 2

α	2	5	8	10
$\zeta(\alpha)$	1.645	1.037	1.004	1.001
Bound (2.8)	1.157	1.007	1.000	1.000

2.2.2. Exponentials of Primes

A classic result in number theory is that the sum of the reciprocals of primes diverges. It is clear that $\sum_p e^{-\lambda p}$ converges for $\lambda > 0$. Accurate analytic bounds on this sum could be of interest since there is no analytic formula for this convergent sum. One such bound is obtained below by using part b) of our Theorem 1.

Towards this end, choose $f(x)$ to be the exponential density $\lambda e^{-\lambda x}$. Then, by part b) of Theorem 1, for all x and $r > 0$,

$$\begin{aligned} \frac{1}{x^r} &\geq \frac{r\lambda^x e^{r+1}}{(r+1)^{r+1}} e^{-\lambda x} \\ \Rightarrow 1 - \frac{1}{x^r} &\leq 1 - \frac{r\lambda^r\ e^{r+1}}{(r+1)^{r+1}}\ e^{-\lambda x}. \end{aligned} \tag{2.9}$$

If we now use the **Euclid identity**

$$\zeta(r) = \prod_p \frac{1}{1 - \frac{1}{p^r}}, \quad \text{(for } r > 1) \tag{2.10}$$

Then from (2.9) we obtain

$$\begin{aligned} \zeta(r) &\geq \prod_p \left(\frac{1}{1 - \frac{r\lambda^r\ e^{r+1}\ e^{-\lambda p}}{(r+1)^{r+1}}}\right) \\ &\geq \prod_p \left(1 + \frac{r\lambda^r e^{r+1}}{(r+1)^{r+1}} e^{-\lambda p}\right) \\ &\geq 1 + \frac{r\lambda^r e^{r+1}}{(r+1)^{r+1}} \sum_p e^{-\lambda p}, \end{aligned}$$

resulting in the analytic upper bound

$$\forall r > 1, \sum_p e^{-\lambda p} \leq \frac{(r+1)^{r+1}}{r\lambda^r e^{r+1}}(\zeta(r) - 1). \tag{2.11}$$

By choosing r appropriately in (2.11), reasonable sharp bounds are obtained. We will not report them here, because we wish to emphasize the analytic nature of the (2.11) bound.

2.3. CDF vs. Density.

The normal distribution $N(\mu, \sigma^2)$ has the very useful property $P(|X - \mu| > k\sigma) \leq \frac{2\phi(k)}{k}$ for $k > 0$. Such bounds are useful because the tail probability is usually more clumsy than the density as a function of k. It is a pleasant consequence of part b) of Theorem 1 that we can establish such bounds in much greater generality; specific examples are given following the next result.

2.3.1 A Tail Bound.

Proposition 2.

Let X have the CDF $F(x)$ and let $|X|$ have the density function $f(x)$. Suppose f is one differentiable and $\psi(x) = -\frac{xf'(x)}{f(x)}$ is increasing for $x > 0$. Then for any k such that $\psi(k) > 1$,

$$P(|X| > k) \leq \frac{k}{\psi(k) - 1} f(k) \tag{2.12}$$

Remark.

Suppose X has a symmetric log concave density; the normal and the double exponential are of this kind (although t densities are not). Then $\psi(x)$ is automatically increasing for $x > 0$, being the product of two positive increasing functions and so the bound (2.12) applies. We will shortly see that bound (2.12) can hold even absent log concavity.

Proof of Proposition 2:

From part b) of Theorem 1, $$P(|X| > k) \leq \frac{\sup_{x>0}\{x^{r+1}f(x)\}}{rk^r}, \tag{2.13}$$

for every $r > 0$.

Now $x^{r+1}f(x)$ is maximized when $\psi(x) = r + 1$ (note there is at most one such x if $\psi(\cdot)$ is increasing). The trick is to choose r appropriately. Choose r to be such that this x is equal to the given k, i.e., $r = \psi(k) - 1$.

Then from (2.13),

$$P(|X| > k) \le \frac{k^{r+1} f(k)}{r k^r} = \frac{k f(k)}{\psi(k) - 1},$$

which proves Proposition 2.

Corollary 1.

From (2.12), on lengthy but straightforward calculations, one gets the following nice bounds:

a) Let X have the t density $f_\alpha(x)$ with α degrees of freedom.

Then for all $k > 1$,

$$P(X > k) \le \frac{k(\alpha + k^2)}{\alpha(k^2 - 1)} f_\alpha(k) \tag{2.14}$$

b) Let X have the gamma density $f_{\alpha,\lambda}(x) = e^{-\lambda x} x^{\alpha-1} \frac{\lambda^\alpha}{\Gamma(\alpha)}$.

Then for all $k > E(X) = \frac{\alpha}{\lambda}$,

$$P(X > k) \le \frac{k}{k\lambda - \alpha} f_{\alpha,\lambda}(k) \tag{2.15}$$

c) Let X have the F density $f_{\alpha,\beta}(x)$ with degrees of freedom α and β. Then for all $k > 1$,

$$P(X > k) \le \frac{2k(\beta + \alpha k)}{\alpha\beta(k - 1)} f_{\alpha,\beta}(k) \tag{2.16}$$

d) Let X have the Beta (α, β) density $f_{\alpha,\beta}(x)$. If $\beta > 1$, then for all $k > \frac{\alpha}{\alpha+\beta-1}$,

$$P(X > k) \le \frac{k(1 - k)}{k(\alpha + \beta - 1) - \alpha} f_{\alpha,\beta}(k) \tag{2.17}$$

2.3.2 Bound on P-values.

The tail bounds in Corollary 1 are specifically useful in finding bounds on P-values for some common tests, e.g., the t, F, and chi-square tests. Note that unlike the normal table, statistics texts never give tables of CDFs for these distributions and some software packages, e.g. SAS, report all P-values below .001 as .001. Analytical bounds such as those in Corollary 1, if they could be made better known, may be useful in some applications.

As a case in point, consider computation of a two-sided P-value for the student t test. The following table illustrates the bound obtained from equation (2.14) when the obtained value of the t statistic is 3 and 4, respectively.

TABLE 3

Degree of Freedom	10	15	20	25	30
True P-value	.013	.009	.007	.006	.0054
Bound (2.14)	.016	.011	.009	.007	.0066
True P-value	.0025	.001	.0007	.0005	.0004
Bound (2.14)	.0028	.001	.0008	.0006	.0004

3. Application to Multivariate Probabilities.

The result of Theorem 1 clearly can be applied to univariate functions of a multivariate random vector. Important special classes in the multivariate case are the spherically symmetric unimodal and normal scale mixture classes. The latter is a subclass of the spherically symmetric unimodal family. We will characterize the best constant $C^*(r)$ of Theorem 1 for each of these families when the event in consideration is of the form $A = \{\underset{\sim}{X} : ||\underset{\sim}{X}|| > k\}$, where $|| \cdot ||$ denotes L_2 norm. It will be seen that $C^*(r)$ is significantly smaller for the normal scale mixtures. The following lemma is needed.

Lemma 1. Let $\underset{\sim}{Z}_{p\times 1}$ be any random vector and let $C^*_{\underset{\sim}{Z}}(r)$ have the property

$$P(||\underset{\sim}{Z}|| > k) \leq C^*_{\underset{\sim}{Z}}(r)\frac{E||\underset{\sim}{Z}||^r}{k^r}$$

for all $k > 0$. Let $\sigma \geq 0$ be any nonnegative random variable independent of $\underset{\sim}{Z}$ and let $\underset{\sim}{X} = \sigma\underset{\sim}{Z}$. Then $P(||\underset{\sim}{X}|| > k) \leq C^*_{\underset{\sim}{Z}}(r)\frac{E||\underset{\sim}{X}||^r}{k^r}$ for all $k > 0$. Hence, in particular, $C^*_{\underset{\sim}{X}}(r) \leq C^*_{\underset{\sim}{Z}}(r)$.

The proof of Lemma 1 is clear on using the fact $E||\underset{\sim}{X}||^r = E(\sigma^r)E(||\underset{\sim}{Z}||^r)$.

Theorem 3.

a) Let $\underset{\sim}{X}$ have a multivariate normal scale mixture distribution, i.e., $\underset{\sim}{X}$ has a density which is a mixture of $N_p(\underset{\sim}{0}, \sigma^2 I)$ densities. Suppose $f_p(x), F_p(x)$ denote respectively the *pdf* and the *CDF* of a $\chi^2(p)$ distribution and $x_* = x_*(p, r)$ denotes the <u>*unique*</u>

root of the equation $xf_p(x) = \frac{r}{2}(1 - F_p(x))$. Then

$$C^*_{\underset{\sim}{X}}(r) \le \frac{2}{r}\frac{\Gamma(\frac{p}{2})}{2^{\frac{r}{2}}\Gamma(\frac{r+p}{2})}x_*^{\frac{r}{2}+1}f_p(x_*) = C^*_{N(\underset{\sim}{0},I)}(r) \tag{3.1}$$

$$\le \frac{2}{r}(\frac{r+p}{2})^{\frac{r+p}{2}}\frac{e^{-\frac{r+p}{2}}}{\Gamma(\frac{r+p}{2})} \overset{def}{=} \overline{C}(r) \tag{3.2}$$

b) Let $\underset{\sim}{X}$ have a spherically symmetric unimodal distribution. Then

$$C^*_{\underset{\sim}{X}}(r) = (\frac{r}{r+p})^{\frac{r}{p}}. \tag{3.3}$$

Proof:

The proof of (3.1) is a repetition of the proof of part a) of Theorem 1 if we note $\underset{\sim}{X} = \sigma\underset{\sim}{Z}$ where $\underset{\sim}{Z} \sim N_p(\underset{\sim}{0}, I), \sigma$ is independent of $\underset{\sim}{Z}$, and so Lemma 1 is applicable. The bound (3.2) follows on an evaluation of $\sup_{x>0}\{x^{\frac{r}{2}+1}f_p(x)\}$ and then some algebra.

For part b), note that $\underset{\sim}{X}$ again has the representation $\underset{\sim}{X} = \sigma\underset{\sim}{Z}$ where $\underset{\sim}{Z}$ is now uniform in the unit ball $B = \{\underset{\sim}{z} : ||\underset{\sim}{z}|| \le 1\}$ and σ is independent of $\underset{\sim}{Z}$. It is easily verified that $P(||\underset{\sim}{Z}|| > z) = \frac{\lambda_{p-1}(B)}{p\lambda_p(B)}(1 - z^p)$ and $E||\underset{\sim}{Z}||^r = \frac{\lambda_{p-1}(B)}{(r+p)\lambda_p(B)}$, where $\lambda_p(B)$ is the volume and $\lambda_{p-1}(B)$ the surface volume of B. Expression (3.3) will therefore follow from the general result (2.1) in Theorem 1.

It is interesting to see the effect of the dimension on the value of the best possible constant $C^*_{\underset{\sim}{X}}(r)$. Here is a brief report. Even in 5 dimensions, $C^*_{\underset{\sim}{X}}(r) \le .44$ for $r = 2$ for the entire normal scale mixture class!

TABLE 4
Normal scale mixture

	$p=2$		$p=3$		$p=4$		$p=5$		$p=6$	
	$C^*_{\underset{\sim}{Z}}(r)$	$\overline{C}(r)$	$C^*_{\underset{\sim}{Z}}(r)$	$\overline{C}(r)$	$C^*_{\underset{\sim}{Z}}(r)$	$\overline{C}(r)$	$C^*_{\underset{\sim}{Z}}(r)$	$\overline{C}(r)$	$C^*_{\underset{\sim}{Z}}(r)$	$\overline{C}(r)$
r										
2	.368	.541	.396	.610	.420	.672	.440	.729	.457	.781
3	.308	.407	.330	.448	.348	.486	.364	.521	.379	.554
4	.271	.336	.287	.364	.302	.391	.315	.415	.328	.439
6	.224	.292	.235	.277	.246	.292	.255	.307	.264	.321
8	.195	.219	.204	.230	.211	.241	.218	.251	.225	.261

TABLE 5: $C^*_X(r)$
Spherically symmetric unimodal

r	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$
2	.5	.543	.577	.606	.630
3	.465	.5	.530	.555	.577
4	.444	.474	.5	.523	.543
6	.422	.444	.465	.483	.5
8	.410	.428	.444	.460	.474

4. The Poisson Case.

The Poisson case, by virtue of well known Poisson approximations in many interesting problems, deserves special mention. The technical aspect is a bit different though, because one is interested in $P(X > k)$ for integer k and so computing $C^*(r)$ involves $\sup_{k\varepsilon Z} k^r P(X > k)$, Z being the positive integers.

To find $\sup_{k\varepsilon Z} k^r P(X > k)$, note that

Step 1.

$$\frac{k^r P(X > k)}{(k-1)^r P(X > k-1)} \geq 1$$

$$\Leftrightarrow \quad \frac{P(X = k)}{P(X > k)} \leq (\frac{k}{k-1})^r - 1 \tag{4.1}$$

Note: we may take $k \geq 2$ in (4.1)

Step 2. By a well known integral representation, if $X \sim Poi(\lambda)$, then $P(X > k) = \frac{\int_0^\lambda e^{-t} t^k \, dt}{k!}$, and so (4.1) is equivalent to

$$\frac{e^{-\lambda}\lambda^k}{\int_0^\lambda e^{-t} t^k dt} \leq (\frac{k}{k-1})^r - 1$$

$$\Leftrightarrow \frac{1}{\int_0^\lambda e^{-t} (\frac{t}{\lambda})^k dt} \leq e^\lambda \left\{ (\frac{k}{k-1})^r - 1 \right\}$$

$$\Leftrightarrow \frac{1}{\int_0^1 e^{-\lambda u} u^k du} \leq \lambda e^\lambda \left\{ \left(\frac{k}{k-1}\right)^r - 1 \right\} \tag{4.2}$$

Step 3. Since $\int_0^1 e^{-\lambda u} u^x du$ is decreasing in x and so is $(\frac{x}{x-1})^r$ for positive r, there is a unique (usually nonintegral) root $x^*(r, \lambda) > 2$ of

$$(\frac{x}{x-1})^r - \frac{e^{-\lambda}\lambda^x}{\int_0^\lambda e^{-t}t^x dt} - 1 = 0, \tag{4.3}$$

and $\arg\sup\{k^r P(X > k)\} = [x^*(r, \lambda)]$, the integer part of $x^*(r, \lambda)$.

One then has the following result, with $S_2(n, i)$ denoting the Stirling numbers of the second kind (see Bryant (1993)), and B_n the nth Bell number.

Theorem 4. Let $X \sim Poi(\lambda)$. Let $k \geq 1$. Then for all $n \geq 1$,

$$P(X > k) \leq C^*(n, \lambda) \frac{\sum_{i=0}^{n} S_2(n, i)\lambda^i}{k^n}, \tag{4.4}$$

where

$$C^*(n, \lambda) = \frac{[x^*(n, \lambda)]^n P(X > [x^*(n, \lambda)])}{B_n} \tag{4.5}$$

Proof: Use the fact that if $X \sim Poi(\lambda)$, then $E(X^n) = \sum_{i=0}^{n} S_2(n, i)\lambda^i$ (see DasGupta (1998) for one reference).

Example 3. Fixed Points in a Random Permutation. Let π_m denote a random permutation of $\{1, 2, \ldots, m\}$ and let Z_m denote the number of its fixed points, i.e., $Z_m = \#\{i : \pi_m(i) = i\}$. It is well known that as $m \to \infty$, $Z_m \xrightarrow{\mathcal{L}} Poi(1)$. Diaconis (1987) gives the total variation bound

$$d_{TV}(Z_m, Poi(1)) \leq \frac{2^m}{m!}. \tag{4.6}$$

If we now use the fact that $\sum_{i=0}^{n} S_2(n, i) = B_n$, the *nth* Bell number, then (4.5) and (4.6) together imply the simple bounds:

$$\text{For every } n, P(Z_m > k) \leq C^*(n, 1)\frac{B_n}{k^n} + \frac{2^m}{m!} \tag{4.7}$$

For instance, for $n = 2, 3, 4, 5, 10$, B_n equals $2, 5, 15, 52, 115975$ and $C^*(n, 1)$ can be computed as $.161, .129, .103, .089, .050$. Thus, for instance, with $n = 5$,

$$P(Z_m > k) \leq \frac{4.628}{k^5} + \frac{2^m}{m!} \tag{4.8}$$

for any m, any $k \geq 1$.

As a numerical test, if $m = 52$ and $k = 3$, the correct value of $P(Z_m > k)$ is .0189882 and $\frac{4.628}{k^5} + \frac{2^m}{m!} = .01904$.

The table of $C^*(n, \lambda)$ below is provided for future use.

TABLE 6: $C^*(n, \lambda)$
Poisson Distribution

n	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 10$
1	.264	.323	.385	.425	.448	.546
2	.161	.216	.265	.297	.320	.399
3	.129	.175	.207	.232	.251	.318
4	.103	.143	.170	.190	.206	.266
5	.089	.119	.141	.159	.174	.229
10	.050	.070	.084	.093	.104	.138

5. Application to Bounding The Mean Absolute Deviation.

The bound

$$P(|X - \mu| > k) \leq C_\mu^*(r) \frac{E|X - \mu|^r}{k^r} \quad \text{where} \quad C_\mu^*(r) = \frac{\sup_{x>0} x^r P(|X - \mu| > x)}{E|X - \mu|^r}$$

is useful to find a family of upper bounds on the mean absolute deviation $E|X - \mu|$ where $\mu = E(X)$. Problems in which the mean absolute deviation arise are plentiful and are beautifully chronicled in Diaconis and Zabell (1991) with much other information. The bounds are given below. Some notation is given first:

$$\left.\begin{array}{l} \text{Let } p = P(X = \mu) (= 0 \text{ if } X \text{ is continuous}) \\ C_\mu^*(r) = \frac{\sup_{x>0} x^r P(|X-\mu|>x)}{E|X-\mu|^r} \\ \alpha(r) = \frac{r}{r-1}(1-p)^{1-\frac{1}{r}} (C_\mu^*(r))^{\frac{1}{r}} \quad (r > 1) \end{array}\right\} \tag{5.1}$$

Theorem 5. For all $r > 1$,

$$E|X - \mu| \leq \alpha(r)(E|X - \mu|^r)^{\frac{1}{r}} \tag{5.2}$$

Remark. (5.2) is of use if and when $\alpha(r) < 1$. We will see illustrations following the proof.

Proof: The main idea is to use the bound $P(|X-\mu| > k) \leq C_\mu^*(r)\frac{E|X-\mu|^r}{k^r}$ with two different values of r for k near 0 and k away from 0.

Step 1. Fix any $\delta > 0$, to be suitably chosen later. We have

$$\begin{aligned} E|X-\mu| &= \int_0^\infty P(|X-\mu| > x)dx \\ &=\leq C_\mu^*(r_1)E|X-\mu|^{r_1}\int_0^\delta \frac{1}{x^{r_1}}dx + C_\mu^*(r_2)E|X-\mu|^{r_2}\int_\delta^\infty \frac{1}{x^{r_2}}dx \\ &\qquad \text{(where } 0 \leq r_1 < 1 < r_2 < \infty) \\ &= \frac{C_\mu^*(r_1)\delta^{1-r_1}}{1-r_1}E|X-\mu|^{r_1} + \frac{C_\mu^*(r_2)\delta^{1-r_2}}{r_2-1}E|X-\mu|^{r_2} \end{aligned} \tag{5.3}$$

In particular, for all $\delta > 0$, all $r_2 > 1$, by taking $r_1 = 0$,

$$E|X-\mu| \leq \delta(1-p) + \frac{C_\mu^*(r_2)\delta^{1-r_2}}{r_2-1}E|X-\mu|^{r_2} \tag{5.4}$$

Step 2. For $a, b > 0$, the function $a\delta + b\delta^{1-r_2}$ is minimized when $\delta = \left(\frac{b(r_2-1)}{a}\right)^{\frac{1}{r_2}}$.

Step 3. Hence, from (5.4),

$$\begin{aligned} E|X-\mu| &\leq \inf_{\delta>0}\left\{\delta(1-p) + \frac{C_\mu^*(r_2)E|X-\mu|^{r_2}}{r_2-1}\delta^{1-r_2}\right\} \\ &= \frac{r_2}{r_2-1}(1-p)^{1-\frac{1}{r_2}}(C_\mu^*(r_2))^{\frac{1}{r_2}}(E|X-\mu|^{r_2})^{\frac{1}{r_2}} \end{aligned} \tag{5.5}$$

on some algebra, as stated in (5.2).

Example 4. Efficacy in Some Common Cases The bound (5.2) is useful when an exact expression for $E|X-\mu|$ is clumsy or absent, but a simple expression for $E|X-\mu|^r$ is available (e.g. if r is an even integer). The following are some illustrative values for some interesting cases; they could also be used by future researchers.

TABLE 7: The Coefficient $\alpha(r)$
Negative Binomial $(K, \frac{1}{2})$ distribution

r	$m = 1$	$m = 2$	$m = 3$	$m = 4$
2	.866	.713	.622	.560
4	.364	.321	.289	.266
6	.231	.210	.430	.443
8	.170	.158	.352	.358

TABLE 8: The Coefficient $\alpha(r)$
Binomial $(n, \frac{1}{2})$ distribution

r	$n = 5$	$n = 10$	$n = 15$
4	1.102	.607	.700
6	.937	.477	1.040
8	.849	1.055	1.011

References

Bryant, V. (1993). Aspects of Combinatorics, Cambridge University Press, New York.

DasGupta, A. (1998). On a differential equation and one step recursion for Poisson moments, Technical Report 98-02, Department of Statistics, Purdue University.

Dharmadhikari, S. and Joag-Dev, K. (1988). Unimodality, Convexity, and Applications, Academic Press, New York.

Diaconis, P. (1987). Method of moments in probability and statistics, AMS Short Course Lecture Notes, Volume 37, AMS, Providence, Rhode Island.

Diaconis, P. and Zabell, S. (1991). Closed form summation for classical distributions, Statistical Science, 6, 284-302.

Sellke, T. M. and Sellke, S. H. (1997). Chebyshev inequalities for unimodal distributions, Amer. Statistician, 51, 1, 34-39.