Best constants in Chebyshev inequalities with various applications*

Anirban DasGupta

Department of Statistics, Purdue University, West Lafayette, IN 47907, USA
(e-mail: dasgupta@stat.purdue.edu)

Received: April 1999

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 \( \sum_{\text{prime } p} e^{-kp} \), 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.

Key words: Camp-Meidell inequality, CDF, Chebyshev inequality, Confidence interval, Density, Mean absolute deviation, Permutations, Poisson, Primes, Spherically symmetric, zeta function

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 \( P(|X| > k) \leq \frac{4 E(X^2)}{9k^2} \) when \( r = 2 \). The Camp-Meidell inequality generalizes

* Research supported by a National Security Agency Grant
Gauss’s result as $P(|X| > k) \leq \left( \frac{r}{r+1} \right)^{E|X|^r} \frac{r}{k^r}$ 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 $E|X - \mu| \leq (E|X - \mu|^r)^{1/r}$ 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 $\mu$. For example, we have the general result that in the entire normal scale mixture family, $X \pm 1.82 \frac{\sigma}{\sqrt{n}}$ is almost a 90% confidence interval for the population mean $\mu$, for every $n$. We also give some analytic bounds on certain number-theoretic functions that follow from our main theorem.

2. 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 $P(|X| > k) \leq C(r) \frac{E|X|^r}{k^r}$ 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 $C^*(r) = C^*(F, r)$. 