Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

Basic tail and concentration bounds

Kaveh S. Nobari

Lectures in High-Dimensional Statistics

Department of Mathematics and Statistics
Lancaster University

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

Contents

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

Motivation

It is often of interest to obtain bounds on the tails of a random variable, or two-sided inequalities, which guarantee that the random variable is close to its mean or median. These slides follow the structure of chapter 2 of Wainwright (2019) to shed light on the elementary techniques for obtaining deviation and concentration inequalities.

One way of controlling a tail probability $P[X \geq t]$ is by controlling the moments of the random variable X, where by controlling higher-order moments of the variable X, we can obtain sharper bounds on tail probabilities. This motivates the "Classical bounds" section of the notes.

We then extend the derivation of bounds to more general functions of the random variables in the "Martingale-based methods" section using martingale decompositions, as opposed to limiting the techniques to deriving bounds on the sum of independent random variables.

Finally, the seminar is concluded with a classical result on the concentration properties of Lipschitz functions of Gaussian variables.

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

From Markov to Chernoff
Sub-Gaussian variables and Hoeffding bounds
Sub-exponential variables and Bernstein bounds
Some one-sided results

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

From Markov to Chernoff
Sub-Gaussian variables and Hoeffding bounds
Sub-exponential variables and Bernstein bounds
Some one-sided results

The most elementary tail bound is Markov's inequality:

Markov's inequality

Given a non-negative random variable X with finite mean - i.e. $\mathbb{E}[X] < \infty$, we have

$$P[X \geq t] \leq \frac{\mathbb{E}[X]}{t}, \quad \forall t > 0$$

It is immediately obvious that Markov's inequality requires only the existence of the first moment. If the random variable X also has finite variance - i.e. $\mathrm{var}(X) < \infty$, we have Chebyshev's inequality:

Chebyshev's inequality

For a random variable X that has a finite mean and variance, we have

$$P[|X - \mu| \geq t] \leq \frac{\mathrm{var}(X)}{t^2}, \quad \forall t > 0$$

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

From Markov to Chernoff
Sub-Gaussian variables and Hoeffding bounds
Sub-exponential variables and Bernstein bounds
Some one-sided results

Proof.

Chebyshev's inequality follows from Markov's inequality, by considering the variable $(X - \mathbb{E}[X])^2$ and the constant t^2. By substituting these in the Markov inequality, we get

$$P[(X - \mathbb{E}[X])^2 \geq t^2] \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} \tag{1}$$

$$P[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} \tag{2}$$

Since, $\mathbb{E}[X] = \mu$ and $\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$, we get

$$P[|X - \mu| \geq t] \leq \frac{\text{var}(X)}{t^2}$$

$\square$

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

From Markov to Chernoff
Sub-Gaussian variables and Hoeffding bounds
Sub-exponential variables and Bernstein bounds
Some one-sided results

The earlier results can be generalised as follows:

Extensions of Markov's inequality

Whenever a variable X has a central moment of order k, an application of Markov's inequality to the random variable $|X - \mu|^k$ yields:

$$P[|X - \mu| \geq t] \leq \frac{\mathbb{E}[|X - \mu|]^k}{t^k}, \quad \forall t > 0.$$

and this is not limited to polynomials $|X - \mu|^k$:

Suppose X has a mgf in a neighbourhood of zero, such that there is a constant $b > 0$ that the functions $\rho(\lambda) = \mathbb{E}[\exp(\lambda(X - \mu))]$ exists for all $\lambda < |b|$. Thus, for any $\lambda \in [0, b]$, we may apply Markov's inequality to the random variable $Y = \exp(\lambda(X - \mu))$, obtaining the upper bound:

$$P[(X - \mu) \geq t] = P[\exp(\lambda(X - \mu)) \geq \exp(\lambda t)] \leq \frac{\mathbb{E}[\exp(\lambda(X - \mu))]}{\exp(\lambda t)}$$

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

From Markov to Chernoff
Sub-Gaussian variables and Hoeffding bounds
Sub-exponential variables and Bernstein bounds
Some one-sided results

By taking the log of both side of the latter inequality, we get:

$$\log P[(X - \mu) \geq t] \leq \log \mathbb{E}[\exp(\lambda(X - \mu))] - \lambda t$$

Optimising, our choice of λ, we can obtain the tightest results that yields the Chernoff bound:

Chernoff bound

$$\log P[(X - \mu) \geq t] \leq \inf_{\lambda \in [0,b]} \{\log \mathbb{E}[\exp(\lambda(X - \mu))] - \lambda t\}.$$

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

From Markov to Chernoff
Sub-Gaussian variables and Hoeffding bounds
Sub-exponential variables and Bernstein bounds
Some one-sided results

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

From Markov to Chernoff
Sub-Gaussian variables and Hoeffding bounds
Sub-exponential variables and Bernstein bounds
Some one-sided results

Evidently, the form of the tail bound obtained using the Chernoff approach depends on the growth rate of the mgf. Naturally, in the study of the tail bounds the random variables are then classified in terms of their mgfs. The simplest type of behaviour is known as sub-Gaussian, which shall be motivated by deriving tail bounds for a Gaussian variables, say, X, such that $X \sim N(\mu, \sigma^2)$, with density

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(X-\mu)^2}{2\sigma^2}\right)$$

and thus, the mgf

$$\mathbb{E}[\exp(\lambda X)] = \exp\left(\mu\lambda + \frac{\sigma^2\lambda^2}{2}\right), \quad \forall \lambda \in \mathbb{R}$$

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

From Markov to Chernoff
Sub-Gaussian variables and Hoeffding bounds
Sub-exponential variables and Bernstein bounds
Some one-sided results

Example (Gaussian tail bounds)

Let $X \sim N(\mu, \sigma^2)$ be a Gaussian r.v., which has mgf

$$\mathbb{E}[\exp(\lambda X)] = \exp\left(\mu\lambda + \frac{\sigma^2\lambda^2}{2}\right), \quad \forall \lambda \in \mathbb{R}$$

substituting this into the optimising problem of the Chernoff bound, we get

$$\inf_{\lambda \geq 0}\{\log \mathbb{E}[\exp(\lambda(X - \mu)) - \lambda t]\} = \inf_{\lambda \geq 0}\left\{\frac{\sigma^2\lambda^2}{2} - \lambda t\right\} = -\frac{t^2}{2\sigma^2} \quad (3)$$

Therefore, we can conclude that any $N(\mu, \sigma^2)$ r.v. satisfies the upper deviation inequality

$$P[X \geq \mu + t] \leq \exp\left(-\frac{t^2}{2\sigma^2}\right)$$

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

From Markov to Chernoff
Sub-Gaussian variables and Hoeffding bounds
Sub-exponential variables and Bernstein bounds
Some one-sided results

Proof of equation (3).

To solve the optimisation problem below

$$\inf_{\lambda \geq 0} \left\{ \frac{\sigma^2 \lambda^2}{2} - \lambda t \right\}$$

we take derivatives to find the optimum of this quadratic function, - i.e.

$$\frac{\partial}{\partial \lambda} \left(\frac{\sigma^2 \lambda^2}{2} - \lambda t \right) = 0,$$

which leads to $\lambda_{opt} = \frac{t}{\sigma^2}$. Substituting λ_{opt} with λ in the above equation yields relationship (3). □

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

From Markov to Chernoff
Sub-Gaussian variables and Hoeffding bounds
Sub-exponential variables and Bernstein bounds
Some one-sided results

Definition (Sub-Gaussianity)

A r.v. X with mean $\mu = \mathbb{E}[X]$ is sub-Gaussian if there is a positive number σ, such that

$$\mathbb{E}[\exp(\lambda(X - \mu))] \leq \exp\left(\frac{\sigma^2\lambda^2}{2}\right), \quad \forall \lambda \in \mathbb{R}$$

where the constant σ is referred to as the sub-Gaussian parameter. Moreover, by the symmetry of the definition, the variable $-X$ is sub-Gaussian iff X is sub-Gaussian, so that we also have lower deviation inequality $P[X \leq \mu - t] \leq \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad \forall t \geq 0$. Thus, we conclude that any sub-Gaussian variable satisfies the concentration inequality

$$P[|X - \mu| \geq t] \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right), \quad \forall t \in \mathbb{R}.$$

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

From Markov to Chernoff
Sub-Gaussian variables and Hoeffding bounds
Sub-exponential variables and Bernstein bounds
Some one-sided results

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

From Markov to Chernoff
Sub-Gaussian variables and Hoeffding bounds
Sub-exponential variables and Bernstein bounds
Some one-sided results

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

Classical bounds
Martingale-based methods
Lipschitz functions of Gaussian variables
References

References

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.