CONVEX OPTIMIZATION PROBLEMS

MAT 220

Dr.C. Rajan (8113053359) Off: S109E

INEQUALITIES IN STATISTICS

There are hundreds of inequalities in statistics

See

https://en.wikipedia.org/wiki/List_of_inequalities

for more than 200 inequalities in probability, statistics, number theory, and related fields

In statistics, inequalities play a crucial role, particularly when discussing probability distributions, confidence intervals, and hypothesis testing. Here are some common types of inequalities used in statistics:

1. **Inequality in Probability:**
   - For any event A, the probability is always between 0 and 1, inclusive:
     $0 \leq P(A) \leq 1$
2. **Chebyshev's Inequality:**
   - Chebyshev's Inequality provides a bound on the probability that a random variable deviates from its mean by more than a certain number of standard deviations. For any random variable X and any k > 0:
     $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$
   - This inequality is useful for understanding the dispersion of data in a distribution.
3. **Markov's Inequality:**
   - Markov's Inequality provides an upper bound on the probability that a non-negative random variable is at least a certain value. For any random variable X and any $a > 0$:
     $P(X \geq a) \leq \frac{E(X)}{a}$
   - It is particularly useful when dealing with non-negative variables.
4. **Jensen's Inequality:**
   - Jensen's Inequality deals with the expectation of convex or concave functions of a random variable. If f is a convex function and X is a random variable, then:
     $E[f(X)] \geq f(E[X])$
   - If f is concave, the inequality flips.
5. **Cauchy-Schwarz Inequality:**
   - The Cauchy-Schwarz Inequality is often used in statistics to prove other inequalities or to establish bounds on statistical measures. For two random variables X and Y:
     $|E[XY]| \leq \sqrt{E[X^2] \cdot E[Y^2]}$

AM,GM,HM INEQUALITY

AM = Arithmetic Mean, GM=Geometric Mean, HM=Harmonic Mean

The inequality states that **AM >= GM >= HM** for any sample with n>=2

The result is easy to prove when $n = 2$. In this case we need to show,

$$\frac{a_1 + a_2}{2} \geq \sqrt{a_1 a_2} \geq \frac{2}{\frac{1}{a_1} + \frac{1}{a_2}}. \tag{5}$$

Observe that

$$\frac{a_1 + a_2}{2} - \sqrt{a_1 a_2} = \frac{(\sqrt{a_1} - \sqrt{a_2})^2}{2} \geq 0. \tag{6}$$

This proves that $A_2 \geq G_2$. To prove $G_2 \geq H_2$ observe that writing $b_1 = 1/a_1$ and $b_2 = 1/a_2$ reduces the relevant inequality to

$$\frac{b_1 + b_2}{2} \geq \sqrt{b_1 b_2} \tag{7}$$

which we have already proved. This proves the AM–GM–HM inequality for $n = 2$. Note that to prove $G_n \geq H_n$ for any n is same as proving $A_n \geq G_n$ for the reciprocals of the given real

Restating the inequality using mathematical notation, we have that for any list of n nonnegative real numbers $x_1, x_2, \ldots, x_n$,

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdots x_n},$$

and that equality holds if and only if $x_1 = x_2 = \cdots = x_n$.

Geometric interpretation [edit]

In two dimensions, $2x_1 + 2x_2$ is the perimeter of a rectangle with sides of length x_1 and x_2. Similarly, $4\sqrt{x_1 x_2}$ is the perimeter of a square with the same area, $x_1 x_2$, as that rectangle. Thus for $n = 2$ the AM–GM inequality states that a rectangle of a given area has the smallest perimeter if that rectangle is also a square.

The full inequality is an extension of this idea to n dimensions. Consider an n-dimensional box with edge lengths $x_1, x_2, \ldots, x_n$. Every vertex of the box is connected to n edges of different directions, so the average length of edges incident to the vertex is $(x_1 + x_2 + \cdots + x_n)/n$. On the other hand, $\sqrt[n]{x_1 x_2 \cdots x_n}$ is the edge length of an n-dimensional cube of equal volume, which therefore is also the average length of edges incident to a vertex of the cube.

Thus the AM–GM inequality states that only the n-cube has the smallest average length of edges connected to each vertex amongst all n-dimensional boxes with the same volume.[3]

Examples [edit]

Example 1 [edit]

If $a, b, c > 0$, then the A.M.-G.M. tells us that

$$(1+a)(1+b)(1+c) \geq 2\sqrt{1 \cdot a} \cdot 2\sqrt{1 \cdot b} \cdot 2\sqrt{1 \cdot c} = 8\sqrt{abc}$$

Chebyshev's Inequality:

Let X be any random variable. If you define $Y = (X - EX)^2$, then Y is a nonnegative random variable, so we can apply Markov's inequality to Y. In particular, for any positive real number b, we have

$$P(Y \geq b^2) \leq \frac{EY}{b^2}.$$

But note that

$$EY = E(X - EX)^2 = Var(X),$$
$$P(Y \geq b^2) = P\big((X - EX)^2 \geq b^2\big) = P\big(|X - EX| \geq b\big).$$

Thus, we conclude that

$$P\big(|X - EX| \geq b\big) \leq \frac{Var(X)}{b^2}.$$

This is **Chebyshev's inequality**.

Chebyshev's Inequality

If X is any random variable, then for any $b > 0$ we have

$$P\big(|X - EX| \geq b\big) \leq \frac{Var(X)}{b^2}.$$

Chebyshev's inequality states that the difference between X and EX is somehow limited by $Var(X)$. This is intuitively expected as variance shows on average how far we are from the mean.

Let $X \sim Binomial(n,p)$. Using Chebyshev's inequality, find an upper bound on $P(X \geq \alpha n)$, where $p < \alpha < 1$. Evaluate the bound for $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$.

Solution

One way to obtain a bound is to write

$$\begin{aligned} P(X \geq \alpha n) &= P(X - np \geq \alpha n - np) \\ &\leq P(|X - np| \geq n\alpha - np) \\ &\leq \frac{Var(X)}{(n\alpha - np)^2} \\ &= \frac{p(1-p)}{n(\alpha - p)^2}. \end{aligned}$$

For $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$, we obtain

$$P(X \geq \frac{3n}{4}) \leq \frac{4}{n}.$$

6.2.2 Markov and Chebyshev Inequalities

Let X be any positive continuous random variable, we can write

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= \int_{0}^{\infty} x f_X(x)dx && \text{(since } X \text{ is positive-valued)} \\ &\geq \int_{a}^{\infty} x f_X(x)dx && \text{(for any } a > 0) \\ &\geq \int_{a}^{\infty} a f_X(x)dx && \text{(since } x > a \text{ in the integrated region)} \\ &= a\int_{a}^{\infty} f_X(x)dx \\ &= aP(X \geq a). \end{aligned}$$

Thus, we conclude

$$P(X \geq a) \leq \frac{EX}{a}, \qquad \text{for any } a > 0.$$

We can prove the above inequality for discrete or mixed random variables similarly (using the generalized PDF), so we have the following result, called **Markov's inequality**.

Markov's Inequality

If X is any nonnegative random variable, then

$$P(X \geq a) \leq \frac{EX}{a}, \qquad \text{for any } a > 0.$$

Let $X \sim Binomial(n, p)$. Using Markov's inequality, find an upper bound on $P(X \geq \alpha n)$, where $p < \alpha < 1$. Evaluate the bound for $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$.

Solution

Note that X is a nonnegative random variable and $EX = np$. Applying Markov's inequality, we obtain

$$P(X \geq \alpha n) \leq \frac{EX}{\alpha n} = \frac{pn}{\alpha n} = \frac{p}{\alpha}.$$

For $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$, we obtain

$$P(X \geq \frac{3n}{4}) \leq \frac{2}{3}.$$

6.2.4 Cauchy-Schwarz Inequality

You might have seen the **Cauchy-Schwarz inequality** in your linear algebra course. The same inequality is valid for random variables. Let us state and prove the Cauchy-Schwarz inequality for random variables.

> Cauchy-Schwarz Inequality
>
> For any two random variables X and Y, we have
>
> $$|EXY| \leq \sqrt{E[X^2]E[Y^2]},$$
>
> where equality holds if and only if $X = \alpha Y$, for some constant $\alpha \in \mathbb{R}$.

You can prove the Cauchy-Schwarz inequality with the same methods that we used to prove $|\rho(X,Y)| \leq 1$ in Section 5.3.1. Here we provide another proof. Define the random variable $W = (X - \alpha Y)^2$. Clearly, W is a nonnegative random variable for any value of $\alpha \in \mathbb{R}$. Thus, we obtain

$$\begin{aligned} 0 \leq EW &= E(X - \alpha Y)^2 \\ &= E[X^2 - 2\alpha XY + \alpha^2 Y^2] \\ &= E[X^2] - 2\alpha E[XY] + \alpha^2 E[Y^2]. \end{aligned}$$

So, if we let $f(\alpha) = E[X^2] - 2\alpha E[XY] + \alpha^2 E[Y^2]$, then we know that $f(\alpha) \geq 0$, for all $\alpha \in \mathbb{R}$. Moreover, if $f(\alpha) = 0$ for some α, then we have $EW = E(X - \alpha Y)^2 = 0$, which essentially means $X = \alpha Y$ with probability one. To prove the Cauchy-Schwarz inequality, choose $\alpha = \frac{EXY}{EY^2}$. We obtain

$$\begin{aligned} 0 &\leq E[X^2] - 2\alpha E[XY] + \alpha^2 E[Y^2] \\ &= E[X^2] - 2\frac{EXY}{EY^2} E[XY] + \frac{(EXY)^2}{(EY^2)^2} E[Y^2] \\ &= E[X^2] - \frac{(E[XY])^2}{EY^2}. \end{aligned}$$

Thus, we conclude

$$(E[XY])^2 \leq E[X^2]E[Y^2],$$

which implies

$$|EXY| \leq \sqrt{E[X^2]E[Y^2]}.$$

Using the Cauchy-Schwarz inequality, show that for any two random variables X and Y

$$|\rho(X, Y)| \leq 1.$$

Also, $|\rho(X, Y)| = 1$ if and only if $Y = aX + b$ for some constants $a, b \in \mathbb{R}$.

Solution

Let

$$U = \frac{X - EX}{\sigma_X}, \qquad V = \frac{Y - EY}{\sigma_Y}.$$

Then $EU = EV = 0$, and $Var(U) = Var(V) = 1$. Using the Cauchy-Schwarz inequality for U and V, we obtain

$$|EUV| \leq \sqrt{E[U^2]E[V^2]} = 1.$$

But note that $EUV = \rho(X, Y)$, thus we conclude

$$|\rho(X, Y)| \leq 1,$$

where equality holds if and only if $V = \alpha U$ for some constant $\alpha \in \mathbb{R}$. That is

$$\frac{Y - EY}{\sigma_Y} = \alpha \frac{X - EX}{\sigma_X},$$

which implies

$$Y = \frac{\alpha \sigma_Y}{\sigma_X} X + \left(EY - \frac{\alpha \sigma_Y}{\sigma_X} EX\right).$$

In the Solved Problems section, we provide a generalization of the Cauchy-Schwrarz inequality, called *Hölder's inequality*.

6.2.5 Jensen's Inequality

Remember that variance of every random variable X is a positive value, i.e.,

$$Var(X) = EX^2 - (EX)^2 \geq 0.$$

Thus,

$$EX^2 \geq (EX)^2.$$

If we define $g(x) = x^2$, we can write the above inequality as

$$E[g(X)] \geq g(E[X]).$$

The function $g(x) = x^2$ is an example of **convex** function. **Jensen's inequality** states that, for any convex function g, we have $E[g(X)] \geq g(E[X])$. So what is a convex function? Figure 6.2 depicts a convex function. A function is convex if, when you pick any two points on the graph of the function and draw a line segment between the two points, the entire segment lies above the graph. On the other hand, if the line segment always lies below the graph, the function is said to be **concave**. In other words, $g(x)$ is convex if and only if $-g(x)$ is concave.

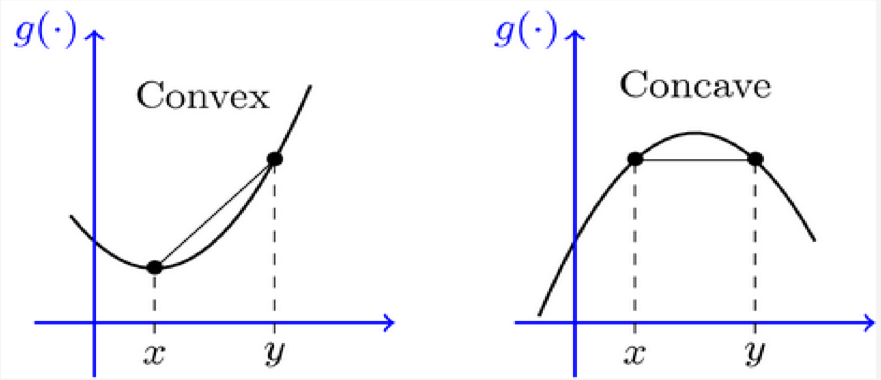

Fig.6.2 - Pictorial representation of a convex function and a concave function.

We can state the definition for convex and concave functions in the following way:

HTTPS://EN.WIKIPEDIA.ORG/WIKI/JENSEN'S_INEQUALITY

The classical form of Jensen's inequality involves several numbers and weights. The inequality can be stated quite generally using either the language of measure theory or (equivalently) probability. In the probabilistic setting, the inequality can be further generalized to its *full strength*.

Finite form [edit]

For a real convex function φ, numbers $x_1, x_2, \ldots, x_n$ in its domain, and positive weights a_i, Jensen's inequality can be stated as:

$$\varphi\left(\frac{\sum a_i x_i}{\sum a_i}\right) \le \frac{\sum a_i \varphi(x_i)}{\sum a_i} \qquad \textbf{(1)}$$

and the inequality is reversed if φ is concave, which is

$$\varphi\left(\frac{\sum a_i x_i}{\sum a_i}\right) \ge \frac{\sum a_i \varphi(x_i)}{\sum a_i}. \qquad \textbf{(2)}$$

Equality holds if and only if $x_1 = x_2 = \cdots = x_n$ or φ is linear on a domain containing $x_1, x_2, \cdots, x_n$.

As a particular case, if the weights a_i are all equal, then (1) and (2) become

$$\varphi\left(\frac{\sum x_i}{n}\right) \le \frac{\sum \varphi(x_i)}{n} \qquad \textbf{(3)}$$

$$\varphi\left(\frac{\sum x_i}{n}\right) \ge \frac{\sum \varphi(x_i)}{n} \qquad \textbf{(4)}$$

For instance, the function $\log(x)$ is *concave*, so substituting $\varphi(x) = \log(x)$ in the previous formula (4) establishes the (logarithm of the) familiar arithmetic-mean/geometric-mean inequality:

$$\log\left(\frac{\sum_{i=1}^n x_i}{n}\right) \ge \frac{\sum_{i=1}^n \log(x_i)}{n} \quad \text{or} \quad \frac{x_1 + x_2 + \cdots + x_n}{n} \ge \sqrt[n]{x_1 \cdot x_2 \cdots x_n}$$

A common application has x as a function of another variable (or set of variables) t, that is, $x_i = g(t_i)$. All of this carries directly over to the general continuous case: the weights a_i are replaced by a non-negative integrable function $f(x)$, such as a probability distribution, and the summations are replaced by integrals.

Definition 6.3

Consider a function $g: I \to \mathbb{R}$, where I is an interval in $\mathbb{R}$. We say that g is a **convex** function if, for any two points x and y in I and any $\alpha \in [0,1]$, we have

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha)g(y).$$

We say that g is **concave** if

$$g(\alpha x + (1-\alpha)y) \geq \alpha g(x) + (1-\alpha)g(y).$$

Note that in the above definition the term $\alpha x + (1-\alpha)y$ is the weighted average of x and y. Also, $\alpha g(x) + (1-\alpha)g(y)$ is the weighted average of $g(x)$ and $g(y)$. More generally, for a convex function $g: I \to \mathbb{R}$, and $x_1, x_2, \ldots, x_n$ in I and nonnegative real numbers α_i such that $\alpha_1 + \alpha_2 + \ldots + \alpha_n = 1$, we have

$$g(\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n) \leq \alpha_1 g(x_1) + \alpha_2 g(x_2) + \ldots + \alpha_n g(x_n) \qquad (6.4)$$

If $n = 2$, the above statement is the definition of convex functions. You can extend it to higher values of n by induction.

Now, consider a discrete random variable X with n possible values x_1, $x_2, \ldots, x_n$. In Equation 6.4, we can choose $\alpha_i = P(X = x_i) = P_X(x_i)$. Then, the left-hand side of 6.4 becomes $g(EX)$ and the right-hand side becomes $E[g(X)]$ (by LOTUS). So we can prove the Jensen's inequality in this case. Using limiting arguments, this result can be extended to other types of random variables.

Jensen's Inequality:

If $g(x)$ is a convex function on R_X, and $E[g(X)]$ and $g(E[X])$ are finite, then

$$E[g(X)] \geq g(E[X]).$$

To use Jensen's inequality, we need to determine if a function g is convex. A useful method is the second derivative.

To use Jensen's inequality, we need to determine if a function g is convex. A useful method is the second derivative.

A twice-differentiable function $g : I \to \mathbb{R}$ is convex if and only if $g''(x) \geq 0$ for all $x \in I$.

For example, if $g(x) = x^2$, then $g''(x) = 2 \geq 0$, thus $g(x) = x^2$ is convex over $\mathbb{R}$.

Example 6.24

Let X be a positive random variable. Compare $E[X^a]$ with $(E[X])^a$ for all values of $a \in \mathbb{R}$.

Solution

First note

$$\begin{aligned} E[X^a] &= 1 = (E[X])^a, && \text{if } a = 0, \\ E[X^a] &= EX = (E[X])^a, && \text{if } a = 1. \end{aligned}$$

So let's assume $a \neq 0, 1$. Letting $g(x) = x^a$, we have

$$g''(x) = a(a-1)x^{a-2}.$$

On $(0, \infty)$, we can say $g''(x)$ is positive, if $a < 0$ or $a > 1$. It is negative, if $0 < a < 1$. Therefore we conclude that $g(x)$ is convex, if $a < 0$ or $a > 1$. It is concave, if $0 < a < 1$. Using Jensen's inequality we conclude

$$\begin{aligned} E[X^a] &\geq (E[X])^a, && \text{if } a < 0 \text{ or } a > 1, \\ E[X^a] &\leq (E[X])^a, && \text{if } 0 < a < 1. \end{aligned}$$

Let $X \sim Exponential(\lambda)$. Using Markov's inequality find an upper bound for $P(X \geq a)$, where $a > 0$. Compare the upper bound with the actual value of $P(X \geq a)$.

Solution

If $X \sim Exponential(\lambda)$, then $EX = \frac{1}{\lambda}$, using Markov's inequality

$$P(X \geq a) \leq \frac{EX}{a} = \frac{1}{\lambda a}.$$

The actual value of $P(X \geq a)$ is $e^{-\lambda a}$, and we always have $\frac{1}{\lambda a} \geq e^{-\lambda a}$.

Let $X \sim Exponential(\lambda)$. Using Chebyshev's inequality find an upper bound for $P(|X - EX| \geq b)$, where $b > 0$.

Solution

a. We have $EX = \frac{1}{\lambda}$ and $VarX = \frac{1}{\lambda^2}$. Using Chebyshev's inequality, we have

$$\begin{aligned} P(|X - EX| \geq b) &\leq \frac{Var(X)}{b^2} \\ &= \frac{1}{\lambda^2 b^2}. \end{aligned}$$

Let $X \sim Exponential(\lambda)$. Using Chernoff bounds find an upper bound for $P(X \geq a)$, where $a > EX$. Compare the upper bound with the actual value of $P(X \geq a)$.

Solution

If $X \sim Exponential(\lambda)$, then

$$M_X(s) = \frac{\lambda}{\lambda - s}, \quad \text{for} \quad s < \lambda.$$

Using Chernoff bounds, we have

$$\begin{aligned} P(X \geq a) &\leq \min_{s>0} \left[e^{-sa} M_X(s)\right] \\ &= \min_{s>0} \left[e^{-sa} \frac{\lambda}{\lambda - s}\right]. \end{aligned}$$

If $f(s) = e^{-sa} \frac{\lambda}{\lambda - s}$, to find $\min_{s>0} f(s)$ we write

$$\frac{d}{ds} f(s) = 0.$$

Therefore,

$$s^* = \lambda - \frac{1}{a}.$$

Note since $a > EX = \frac{1}{\lambda}$, then $\lambda - \frac{1}{a} > 0$. Thus,

$$P(X \geq a) \leq e^{-s^* a} \frac{\lambda}{\lambda - s^*} = a\lambda e^{1-\lambda a}.$$

The real value of $P(X \geq a)$ is $e^{-\lambda a}$ and we have $e^{-\lambda a} \leq a\lambda e^{1-\lambda a}$, or equivalently, $a\lambda e \geq 1$, which is true since $a > \frac{1}{\lambda}$.