Moment and Probability Inequalities

Statistics 110

Summer 2006

Moment Inequalities

- Schwarz's Inequality (sometimes called Cauchy-Schwarz)

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

Proof. Suppose that $E[X^2] > 0$ and $E[Y^2] > 0$ Let

$$U = \frac{X}{\sqrt{E[X^2]}} \quad \text{and} \quad V = \frac{Y}{\sqrt{E[Y^2]}}$$

It can be shown that $2|UV| \leq U^2 + V^2$. Thus

$$2|E[UV]| \leq 2E[|UV|] \leq E[U^2] + E[V^2] = 2$$

This gives

$$(E[UV])^2 \leq (E[|UV|])^2 \leq 1$$

implying

$$\frac{(E[XY])^2}{E[X^2]E[Y^2]} \leq \frac{(E[|XY|])^2}{E[X^2]E[Y^2]}$$

□

One consequence of this inequality is that $(\mathrm{Cov}(X,Y))^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y)$ or $|\mathrm{Cov}(X,Y)| \leq \sigma_X\sigma_Y$. A consequence of this is that $|\mathrm{Corr}(X,Y)| \leq 1$, a result discussed earlier.

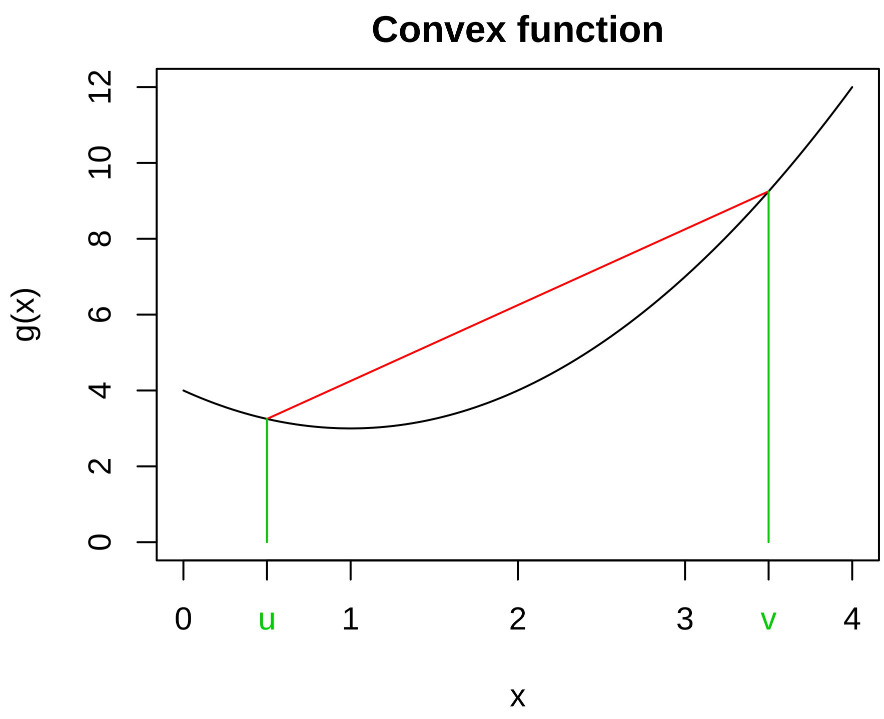
- Jensen's Inequality

  If $g(\cdot)$ is a convex function on the interval (a, b) and X is a RV taking values in (a, b), then $E[g(X)] \geq g(E[X])$.

Note that a function $g(\cdot)$ is convex on the open interval $I = (a, b)$ if

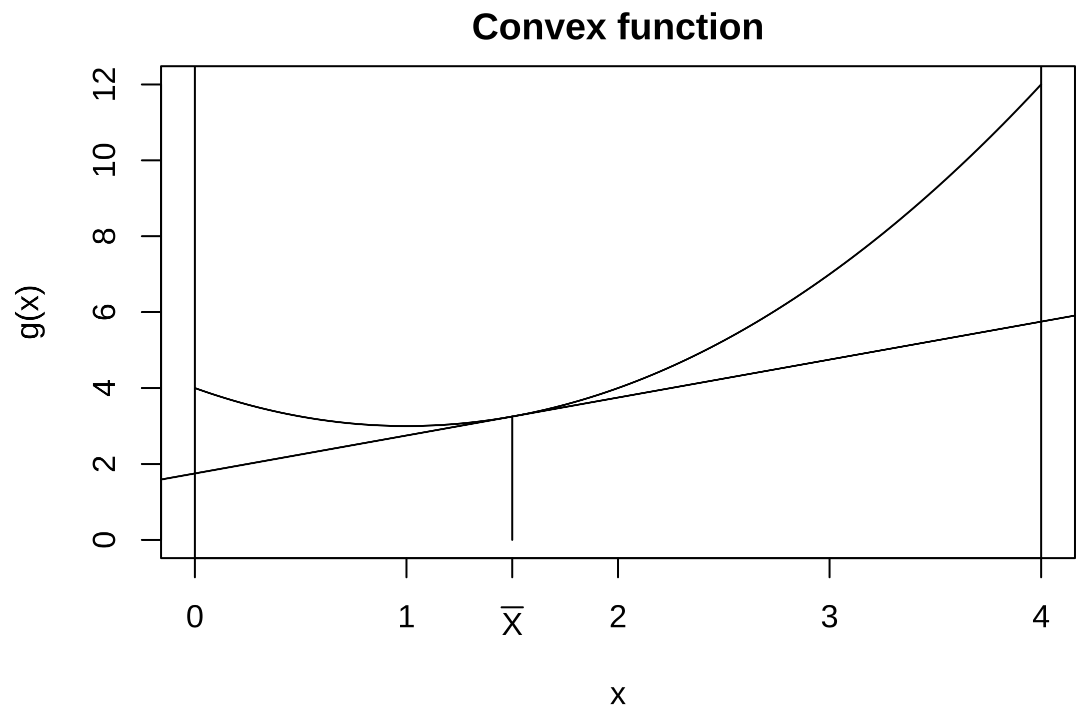
$$g(\alpha u + (1-\alpha)v) \leq \alpha g(u) + (1-\alpha)g(v)$$

for all $u, v \in I$ and $0 \leq \alpha \leq 1$.



Proof. Convexity means that a supporting line exists at each $t \in (a, b)$. i.e. the graph lies completely above each tangent line.

From the supporting line at $t = E[X]$ (with slope λ), we have



$$
\begin{aligned}
g(x) &\geq g(E[X]) + \lambda(x - E[X]) \\
E[g(X)] &\geq E[g(E[X]) + \lambda(X - E[X]) \\
&= g(E[X]) + \lambda(E[X] - E[X]) = g(E[X])
\end{aligned}
$$

□

A couple of examples where Jensen's inequality can be used are the following

1. $E[e^X] \geq \exp(E[X])$.

   For example, assume $X \sim N(\mu, \sigma^2)$ and let $Y = e^X \sim logN(\mu, \sigma^2)$.

   A consequence is that $E[Y] = E[e^X] \geq e^{\mu}$.

   In fact $E[Y] = e^{\mu+0.5\sigma^2}$

   Note going the other way, we get $\log(E[X]) \geq E[\log X]$ since $-\log x$ is a convex function ($\log x$ is a concave function).

   i.e. $\log e^{\mu+0.5\sigma^2} \geq \mu$

2. Arithmetic mean $\geq$ Geometric Mean $\geq$ Harmonic Mean

For any set of n positive numbers $x_1, x_2, \ldots, x_n$,

$$\frac{x_1 + \ldots + x_n}{n} \geq \sqrt[n]{x_1 x_2 \ldots x_n} \geq \frac{n}{\frac{1}{x_1} + \ldots + \frac{1}{x_n}}$$

To justify the first inequality let X be a random variable taking values $x_1, x_2, \ldots, x_n$ each with probability $\frac{1}{n}$. Then Jensen's says

$$\log\left(\frac{x_1 + \ldots + x_n}{n}\right) \geq \frac{\log x_1 + \ldots + \log x_n}{n} = \log(x_1 \ldots x_n)^{1/n}$$

Then exponentiate both sides to get the first inequality.

The other inequalities can be derived similarly.

- Lyapunov's Inequality

  If $0 < s < t$
  $$(E[|X|^s])^{1/s} \le (E[|X|^t])^{1/t}$$

  A consequence of this is the relationship (for some integer p)

  $$E[|X|] \le (E[|X|^2])^{1/2} \le (E[|X|^3])^{1/3} \le \ldots \le (E[|X|^p])^{1/p}$$

  which implies

  $$|E[X]|^q \le (E[|X|])^q \le E[|X|^q] \text{ if } 1 \le q \le p$$

  Proof. Let $r = \frac{t}{s} > 1$. Let $Y = |X|^s$ and apply Jensen's inequality to $g(y) = |y|^r$, giving $(E[|Y|])^r \le E[|Y|^r]$. This implies that

  $$(E[|X|^s])^{t/s} \le E[|X|^t]$$

  Taking the tth root of each side gives the result. $\square$

Probability Inequalities

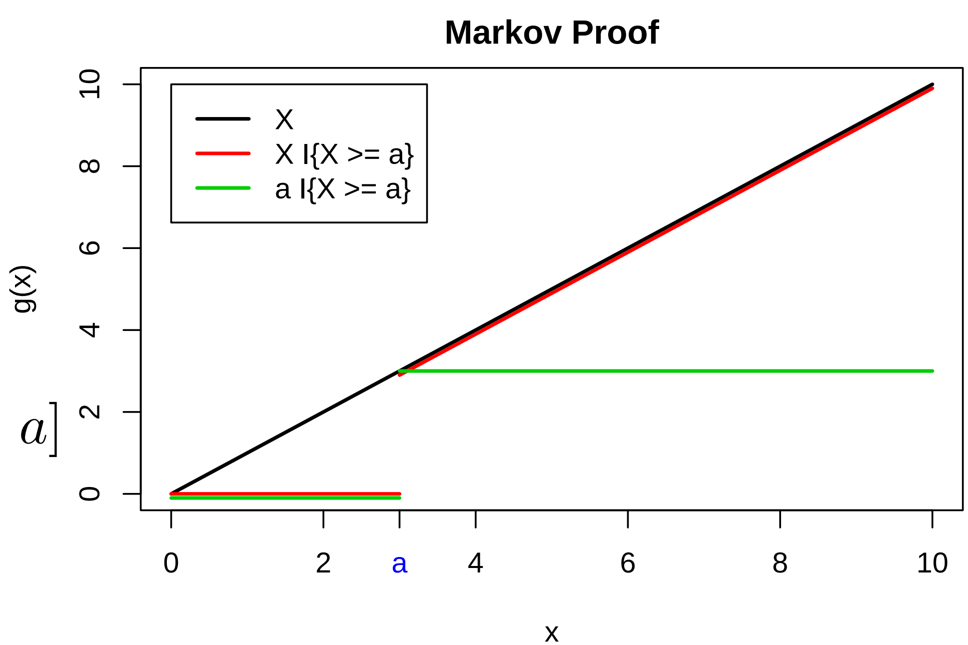
- Markov Inequality

  Let X be a non-negative RV (i.e. $P[X \geq 0] = 1$). Then for any $a > 0$,

$$P[X \geq a] \leq \frac{E[X]}{a}$$

Proof.

$$X \geq XI\{X \geq a\} \geq aI\{X \geq a\}$$



Therefore

$$E[X] \geq E[XI\{X \geq a\}] = aP[X \geq a]$$

□

Note that there is an alternative version of this inequality that says if $[X^r] < \infty$,

$$P[X \geq a] \leq \frac{E[X^r]}{a^r}$$

- Chebyshev's Inequality.

  If $E[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2 < \infty$, then

$$P[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2}$$

  Note that this equality is sometimes written as the equivalent

$$P[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

Proof.

$$P[|X-\mu| \geq a] = P[(X-\mu)^2 \geq a^2] \leq \frac{E[(X-\mu)^2]}{a^2} \quad \text{by Markov's Inequality}$$

Take $a = k$ to get the first form of the result and $a = k\sigma$ to get the second form of the result. $\square$

Example: Suppose it is known that the number of widgets produced for Guinness breweries in a factory during an hour is a RV with mean 500.

1. What can be said about the probability that an hour's production will exceed 1000?
   Answer: By Markov's inequality

$$P[X \geq 1000] \leq \frac{E[X]}{1000} = \frac{500}{1000} = 0.5$$

2. If the variance of a hour's production is known to be 100, then what can be said about the probability that a hour's production will be between 450 and 550?
Answer: By Chebyshev's inequality

$$P[|X - 500| \geq 50] \leq \frac{\mathrm{Var}(X)}{50^2} = \frac{100}{50^2} = \frac{1}{25} = 0.04$$

This implies that

$$P[|X - 500| < 50] \geq 1 - \frac{1}{25} = \frac{24}{25} = 0.96$$

3. What can be said about the probability that the production will be between 450 and 550 if X is normally distributed $(N(500, 100))$?

$$P[450 \leq X \leq 550] = P\left[\frac{450 - 500}{10} \leq Z \leq \frac{550 - 500}{10}\right]$$
$$= P[-5 \leq Z \leq 5] = \Phi(5) - \Phi(-5) = 0.9999994$$

Note that these bounds are not particularly tight in most cases.

In fact they are what happens in a "worst case scenario".

The following inequality also fits into this setting, where the bounds are often loose.

- One-sided Chebyshev's Inequality

  If $E[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2 < \infty$, then for any $a > 0$,

$$P[X \geq \mu + a] \leq \frac{\sigma^2}{\sigma^2 + a^2}$$
$$P[X \leq \mu - a] \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Proof. Without loss of generality, assume that $\mu = 0$. Then for any b,

$$\begin{aligned} P[X \geq a] &= P[X + b \geq a + b] \\ &= P[(X+b)^2 \geq (a+b)^2] \\ &\leq \frac{E[(X+b)^2]}{(a+b)^2} = \frac{E[X^2] + b^2}{(a+b)^2} \\ &= \frac{\alpha + t^2}{(1+t)^2} \stackrel{\text{Def}}{=} g(t) \end{aligned}$$

where

$$\alpha = \frac{E[X^2]}{a^2} = \frac{\sigma^2}{a^2}; \qquad t = \frac{b}{a}$$

To minimize $g(t)$ (i.e. find the best b), set $t = \alpha$, yielding

$$\min g(t) = \frac{\alpha + \alpha^2}{(1+\alpha)^2} = \frac{\frac{\sigma^2}{a^2}}{1 + \frac{\sigma^2}{a^2}} = \frac{\sigma^2}{\sigma^2 + a^2}$$

The other inequality is proved similarly.

□

Example: Back to the widget example. What can be said about the probability that last least 550 widgets are made, assuming the mean is 500 and the variance is 100?

Answer:

$$P[X \geq 550] = P[X \geq 500 + 50] \leq \frac{\sigma^2}{\sigma^2 + 50^2} = \frac{100}{100 + 2500} = 0.0384$$

If we only use the different forms of the Markov inquality we get

$$P[X \geq 550] \leq \frac{E[X]}{550} = \frac{500}{550} = 0.909$$

and

$$P[X \geq 550] \leq \frac{E[X^2]}{550^2} = \frac{\sigma^2 + \mu^2}{550^2} = 0.827$$

Note that if the production was normally distributed, $P[X \geq 550] = 0.000000287$

These probability bounds may not be useful as they may give values greater than 1. For example, if $\mu = 500$, the Markov bound for

$$P[X \geq 400] \leq \frac{500}{400} = 1.25$$

This is a reason why different bounds have been developed. Generally, the stronger the assumptions you make, the tighter the bounds you can get.