Probability Inequalities

October 21, 2008

For a set $A$, let $m_A = \inf\{g(t); t \in A\}$ for a positive function $g$. Then

$$Eg(X) \geq E[g(X)I_A(X)] \geq E[m_AI_A(X)] = m_A P\{X \in A\}.$$  

The Chebyshev inequality occurs by taking $g$ to be function increasing on the support of $X$ and $A = [x, \infty)$, then $m_A = g(x)$,

$$Eg(X) \geq g(x)P\{X > x\} \quad \text{or} \quad P\{X > x\} \leq \frac{Eg(X)}{g(x)}.$$  

This can be seen graphically in Figure 1 for the case $g(x) = x$. The area of the rectangle $xP\{X > x\}$ is less than $EX$, the area above the graph of the cumulative distribution function and below the line $y = 1$.

For the case $X = |Y - \mu_Y|$ and $g(x) = x^2$, we have

$$P\{|Y - \mu_Y| > y\} \leq \frac{E(Y - \mu_Y)^2}{y^2} = \frac{\text{Var}(Y)}{y^2}.$$  

If we choose $g(x) = \exp(tx), t > 0$, then for random variables possessing a moment generating function, the Chebyshev inequality becomes

$$P\{X > x\} \leq \frac{M_X(t)}{\exp(tx)}, \quad \log P\{X > x\} \leq \log M_T(t) - tx.$$  

Next, we minimize this inequality over all possible choices of $t$.

$$\log P\{X > x\} \leq -K^*(x), \quad \inf_{t>0}\{K_T(t) - tx\} = -\sup_{t>0}\{tx - K_T(t)\} = -K^*(x).$$  

where $K_X(t) = \log M_T(x)$, the cumulant generating function, is a convex function. $K^*_X(x)$ is called the (convex) conjugate function for $K_T$ or the rate function. This inequality gives an upper bound for the probability of rare events.

Because $tx - K_T(t)$ is concave down, the maximum of $tx - K_T(t)$ is unique. We take a derivative with respect to $t$ and set the expression equal to 0 to obtain

$$K_X'(t) = x.$$  

(1)

Let $t^*(x)$ denote the solution to Equation (1). Then,

$$K^*_X(x) = t^*(x)x - K_X(t^*(x)).$$
Example 1. For the standard normal, the cumulant generating function, $K_Z(t) = t^2/2$

$$K'_Z(t) = t, \quad t^*(x) = x, \quad K^*(x) = x^2 - \frac{x^2}{2} = \frac{x^2}{2}.$$ 

Thus, for $x > 0$, 

$$P\{Z > x\} \leq \exp\left(-\frac{x^2}{2}\right).$$

The 6σ strategy looks to eliminate errors more common that 6 standard deviations from the mean. For a normal random variables, the rate function tells us that this probability is at most

$$2 \exp(-6^2/2) \approx 3 \times 10^{-8}.$$ 

Example 2. For a Poisson random variable, $M_X(t) = \rho_X(e^t) = \exp(\lambda(e^t - 1))$ and $K_X(t) = \lambda(e^t - 1)$.

$$K'_X(t) = \lambda e^t, \quad t^*(x) = \log \frac{x}{\lambda}, \quad K^*(x) = x \log \frac{x}{\lambda} - x + \lambda.$$ 

Thus, for $x > 0$, 

$$P\{X > x\} \leq \exp -K^*(x) = \left(\frac{\lambda}{x}\right)^x e^{x-\lambda}.$$ 

Exercise 3. For a binomial random variable $M_x(t) = \rho_X(e^t) = ((1-p)+pe^t)^n$. Find the rate function $K^*$. 

---

Figure 1: A geometric proof of the Chebyshev inequality $xP\{X > x\} \leq EX$