Probability Inequalities

October, 2009

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

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

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 $y = 1$.

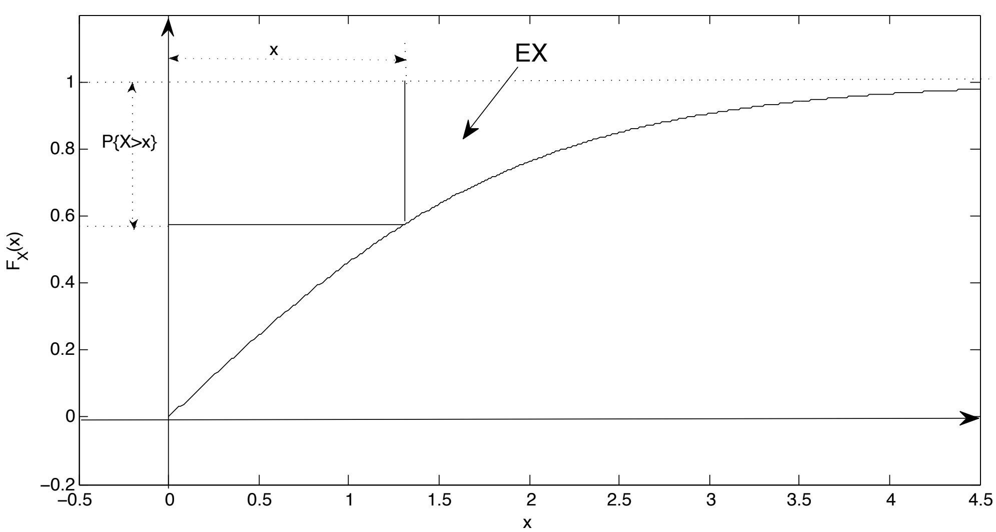

Figure 1: A geometric proof of the Chebyshev inequality $xP\{X > x\} \leq EX$

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

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

Example 1. *For a standard normal random variable,*

$$P\{Z > z\} \leq \frac{Var(Z)}{z^2} = \frac{1}{z^2}.$$

Thus, $P\{Z > 6\} \leq 1/36$.

Can we improve on this estimate?

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.

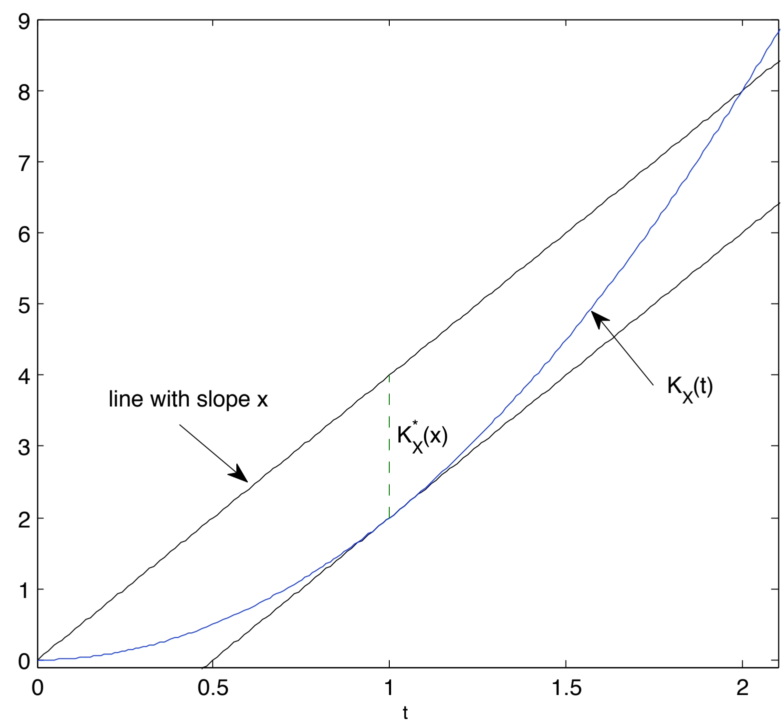

Figure 2: Geometric construction of K_X^*, the rate function

We next show that $K_X(t)$ is a convex function. This will mean that $tx - K_T(t)$ is concave down and so the maximum of $tx - K_T(t)$ is unique. This starts with the following.

Exercise 2. *Let $n : \mathbf{R} \to [0, \infty)$ be a non-negative function with $E_P n(X) > 0$. (Here, we emphasize the probability used in the expectation with the subscript P.) Define, for each event A,*

$$Q(A) = \frac{E_P[I_A n(X)]}{E_P n(X)}.$$

The Q is also a probability.

If we take two derivatives and let $n(x) = \exp tx, t \geq 0$, then

$$K''_X(t) = \frac{M''(t) - M'_X(t)^2}{M_X(t)^2} = \frac{E_P[X^2 e^{tX}]}{E_P e^{tX}} - \left(\frac{E_P[Xe^{tX}]}{E_P e^{tX}}\right)^2 = E_Q X^2 - (E_Q X)^2 = \mathrm{Var}_Q(X) > 0.$$

$K^*_X(x)$ is called the **(convex) conjugate function** or the **Legendre transform** for K_X or the **rate function**. This inequality gives an upper bound for the probability of rare events.

To find the unique maximum of $tx - K_X(t)$, we take a derivative with respect to t and set the expression equal to 0 to obtain

$$K'_X(t) = x. \tag{1}$$

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

$$K^*_X(x) = t^*(x)x - K_X(t^*(x)).$$

Example 3. *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

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

The actual answer is approximately 10^{-9}

Example 4. *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 5. *For a binomial random variable $M_x(t) = \rho_X(e^t) = ((1-p) + pe^t)^n$. Find the rate function K^*.*