556: MATHEMATICAL STATISTICS I

INEQUALITIES

5.1 Concentration and Tail Probability Inequalities

Lemma (CHEBYCHEV'S LEMMA) If X is a random variable, then for non-negative function h, and $c > 0$,

$$\Pr\left[h(X) \geq c\right] \leq \frac{E_{f_X}\left[h(X)\right]}{c}$$

Proof (continuous case) : Suppose that X has density function f_X which is positive for $x \in \mathbb{X}$. Let $\mathcal{A} = \{x \in \mathbb{X} : h(x) \geq c\} \subseteq X$. Then, as $h(x) \geq c$ on $\mathcal{A}$,

$$\begin{aligned} E_{f_X}\left[h(X)\right] = \int h(x) f_X(x)\, dx &= \int_{\mathcal{A}} h(x) f_X(x)\, dx + \int_{\mathcal{A}'} h(x) f_X(x)\, dx \\ &\geq \int_{\mathcal{A}} h(x) f_X(x)\, dx \\ &\geq \int_{\mathcal{A}} c f_X(x)\, dx = c \Pr\left[X \in \mathcal{A}\right] = c \Pr\left[h(X) \geq c\right] \end{aligned}$$

and the result follows.

- **SPECIAL CASE I - THE MARKOV INEQUALITY**
  If $h(x) = |x|^r$ for $r > 0$, so

  $$P\left[|X|^r \geq c\right] \leq \frac{E_{f_X}\left[|X|^r\right]}{c}.$$

  Alternately stated (by Casella and Berger) as follows: If $P[Y \geq 0] = 1$ and $P[Y = 0] < 1$, then for any $r > 0$

  $$P[Y \geq r] \leq \frac{E_{f_X}\left[Y\right]}{r}$$

  with equality if and only if

  $$P[Y = r] = p = 1 - P[Y = 0]$$

  for some $0 < p \leq 1$.

- **SPECIAL CASE II - THE CHEBYCHEV INEQUALITY**
  Suppose that X is a random variable with expectation μ and variance σ^2. Then $h(x) = (x - \mu)^2$ and $c = k^2\sigma^2$, for $k > 0$,

  $$P\left[(X - \mu)^2 \geq k^2\sigma^2\right] \leq 1/k^2$$

  or equivalently

  $$P\left[|X - \mu| \geq k\sigma\right] \leq 1/k^2.$$

  Setting $\epsilon = k\sigma$ gives

  $$P\left[|X - \mu| \geq \epsilon\right] \leq \sigma^2/\epsilon^2$$

  or equivalently

  $$P\left[|X - \mu| < \epsilon\right] \geq 1 - \sigma^2/\epsilon^2.$$

Theorem (CHERNOFF BOUNDS)
Suppose that $X_1, \ldots, X_n$ are independent binary trials (known as "Poisson trials") such that

$$P[X_i = x] = \begin{cases} 1 - p_i & x = 0 \\ p_i & x = 1 \end{cases}$$

and zero otherwise. Let $X = (X_1 + \cdots + X_n)$, so that $E_{f_X}[X] = \sum_{i=1}^n p_i = \mu$, say. Then for $d > 0$

$$P[X \geq (1+d)\mu] \leq \left\{ \frac{e^d}{(1+d)^{(1+d)}} \right\}^\mu .$$

If $0 \leq d \leq 1$, a simpler bound is

$$P[X \geq (1+d)\mu] \leq \exp\left\{-\mu d^2/3\right\}.$$

Proof Let $a > 0$. Then, using the Chebychev Lemma with $h(x) = e^{ax}$, and $c = e^{a(1+d)\mu}$, we have

$$P[X \geq (1+d)\mu] = P[\exp\{aX\} \geq \exp\{a(1+d)\mu\}] \leq \frac{E_{f_X}[\exp\{aX\}]}{\exp\{a(1+d)\mu\}}. \tag{1}$$

But, by independence,

$$E_{f_X}[\exp\{aX\}] \quad = \quad \prod_{i=1}^n E_{f_{X_i}}[\exp\{aX_i\}] = \prod_{i=1}^n [p_i e^a + (1-p_i)] = \prod_{i=1}^n [1 + p_i(e^a - 1)]$$

Now for $y \geq 0$,

$$e^y = 1 + y + \frac{y^2}{2} + \cdots \geq 1 + y$$

so setting $y_i = p_i(e^a - 1)$ and using this inequality term by term, we conclude from equation (1) that

$$E_{f_X}[\exp\{aX\}] = \prod_{i=1}^n [1 + p_i(e^a - 1)] \leq \prod_{i=1}^n \exp\{p_i(e^a - 1)\} = \exp\left\{\sum_{i=1}^n p_i(e^a - 1)\right\} = \exp\left\{\mu(e^a - 1)\right\}.$$

Hence

$$P[X \geq (1+d)\mu] \leq \frac{\exp\left\{\mu(e^a - 1)\right\}}{\exp\{a(1+d)\mu\}}$$

and setting $a = \log(1+d)$ yields

$$P[X \geq (1+d)\mu] \leq \frac{e^{\mu d}}{(1+d)^{\mu(1+d)}} = \left\{ \frac{e^d}{(1+d)^{(1+d)}} \right\}^\mu$$

For $0 \leq d \leq 1$, we have that

$$\left\{ \frac{e^d}{(1+d)^{(1+d)}} \right\}^\mu \leq \exp\{-\mu d^2/3\}.$$

To see this, consider taking logs, and the function

$$g(d) = d - (1+d)\log(1+d) + d^2/3.$$

We need to show that $g(d)$ is bounded above by zero for $0 \leq d \leq 1$. Clearly $g(0) = 0$, and taking derivatives twice we have

$$g^{(1)}(d) = -\log(1+d) + 2d/3 \qquad\qquad g^{(2)}(d) = -\frac{1}{(1+d)} + 2/3.$$

Therefore $g^{(1)}(0) = 0$, $g^{(2)}(0) = -1/3 < 0$ and $g^{(1)}(1) = -\log 2 + 2/3 < 0$, so $g^{(1)}(d)$ stays **negative** for all $0 < d \leq 1$ as there is no solution of $g^{(1)}(d) = 0$ in this interval. Thus $g(d)$ must also be negative for all d in this range.

Theorem (A CHERNOFF BOUND USING MGFS)
If X is a random variable, with mgf $M_X(t)$ defined on a neighbourhood $(-h, h)$ of zero. Then

$$P\left[X \geq a\right] \leq e^{-at} M_X(t) \qquad \text{for } 0 < t < h$$

Proof Using the Chebychev Lemma with $h(x) = e^{tx}$ and $c = e^{at}$, for $t > 0$,

$$P\left[X \geq a\right] = P\left[tX \geq at\right] = P\left[\exp\{tX\} \geq \exp\{at\}\right] \leq \frac{E_{f_X}[e^{tX}]}{e^{at}} = \frac{M_X(t)}{e^{at}}$$

provided $t < h$ also. Using similar methods,

$$P\left[X \leq a\right] \leq e^{-at} M_X(t) \qquad \text{for } -h < t < 0$$

Theorem (TAIL BOUNDS FOR THE NORMAL DENSITY)
If $Z \sim N(0,1)$, then for $t > 0$

$$\sqrt{\frac{2}{\pi}} \frac{t}{1+t^2}\, e^{-t^2/2} \leq \Pr[|Z| \geq t] \leq \sqrt{\frac{2}{\pi}} \frac{1}{t} e^{-t^2/2}$$

Proof By symmetry, $\Pr[|Z| \geq t] = 2\Pr[Z \geq t]$, so

$$P[Z \geq t] \quad = \quad \left(\frac{1}{2\pi}\right)^{1/2} \int_t^\infty e^{-x^2/2}\, dx \leq \left(\frac{1}{2\pi}\right)^{1/2} \int_t^\infty \frac{x}{t} e^{-x^2/2}\, dx = \left(\frac{1}{2\pi}\right)^{1/2} \frac{e^{-t^2/2}}{t}.$$

Similarly, for $t > 0$,

$$\int_t^\infty e^{-x^2/2}\, dx \equiv \int_t^\infty \frac{x}{x}\, e^{-x^2/2}\, dx = \left[-\frac{1}{x} e^{-x^2/2}\right]_t^\infty - \int_t^\infty \frac{1}{x^2}\, e^{-x^2/2}\, dx \geq \frac{1}{t} e^{-t^2/2} - \frac{1}{t^2} \int_t^\infty e^{-x^2/2}\, dx$$

after writing $1 = x/x$, then integrating by parts, and then noting that, on (t, ∞), $x > t \iff 1/x^2 < 1/t^2$, and that the integrand is non-negative. Therefore, combining terms

$$\left(1 + \frac{1}{t^2}\right) \int_t^\infty e^{-x^2/2}\, dx \geq \frac{1}{t}\, e^{-t^2/2}$$

and cross-multiplying by the positive term $t^2/(1+t^2)$ yields

$$\int_t^\infty e^{-x^2/2}\, dx \geq \frac{t}{1+t^2}\, e^{-t^2/2} \qquad \therefore \qquad \Pr[|Z| > t] \geq \sqrt{\frac{2}{\pi}} \frac{t}{1+t^2}\, e^{-t^2/2}.$$

To see the quality of the approximation, the table below shows the values of the bounding values for t ranging from 1 to 5. Clearly the bounds improve as t gets larger.

t	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Lower	2.420e-01	1.196e-01	4.319e-02	1.209e-02	2.659e-03	4.610e-04	6.298e-05	6.770e-06	5.718e-07
True	3.173e-01	1.336e-01	4.550e-02	1.242e-02	2.700e-03	4.653e-04	6.334e-05	6.795e-06	5.733e-07
Upper	4.839e-01	1.727e-01	5.399e-02	1.402e-02	2.955e-03	4.987e-04	6.692e-05	7.104e-06	5.947e-07

5.2 Expectation Inequalities

Lemma Let $a, b > 0$ and $p, q > 1$ satisfy

$$p^{-1} + q^{-1} = 1. \tag{2}$$

Then

$$p^{-1}\, a^p + q^{-1}\, b^q \geq ab$$

with equality if and only if $a^p = b^q$.

Proof Fix $b > 0$. Let

$$g(a; b) = p^{-1}\, a^p + q^{-1}\, b^q - ab.$$

We require that $g(a; b) \geq 0$ for all a. Differentiating wrt a for fixed b yields $g^{(1)}(a; b) = a^{p-1} - b$, so that $g(a; b)$ is minimized (the second derivative is strictly positive at all a) when $a^{p-1} = b$, and at this value of a, the function takes the value

$$p^{-1}\, a^p + q^{-1}\, (a^{p-1})^q - a(a^{p-1}) = p^{-1}\, a^p + q^{-1}\, a^p - a^p = 0$$

as, by equation (2), $1/p + 1/q = 1 \implies (p-1)q = p$. As the second derivative is strictly positive at all a, the minimum is attained at the **unique** value of a where $a^{p-1} = b$, where, raising both sides to power q yields $a^p = b^q$.

Theorem (HÖLDER'S INEQUALITY)
Suppose that X and Y are two random variables, and $p, q > 1$ satisfy 2. Then

$$|E_{f_{X,Y}}[XY]| \leq E_{f_{X,Y}}[|XY|] \leq \{E_{f_X}[|X|^p]\}^{1/p}\, \{E_{f_Y}[|Y|^q]\}^{1/q}$$

Proof (continuous case) For the first inequality,

$$E_{f_{X,Y}}[|XY|] = \iint |xy| f_{X,Y}(x, y)\; dx\; dy \geq \iint xy f_{X,Y}(x, y)\; dx\; dy = E_{f_{X,Y}}[XY]$$

and

$$E_{f_{X,Y}}[XY] = \iint xy f_{X,Y}(x, y)\; dx\; dy \geq \iint -|xy| f_{X,Y}(x, y)\; dx\; dy = -E_{f_{X,Y}}[|XY|]$$

so

$$-E_{f_{X,Y}}[|XY|] \leq E_{f_{X,Y}}[XY] \leq E_{f_{X,Y}}[|XY|] \qquad \therefore \qquad |E_{f_{X,Y}}[XY]| \leq E_{f_{X,Y}}[|XY|].$$

For the second inequality, set

$$a = \frac{|X|}{\{E_{f_X}[|X|^p]\}^{1/p}} \qquad b = \frac{|Y|}{\{E_{f_Y}[|Y|^q]\}^{1/q}}.$$

Then from the previous lemma

$$p^{-1}\, \frac{|X|^p}{E_{f_X}[|X|^p]} + q^{-1}\, \frac{|Y|^q}{E_{f_Y}[|Y|^q]} \geq \frac{|XY|}{\{E_{f_X}[|X|^p]\}^{1/p}\, \{E_{f_Y}[|Y|^q]\}^{1/q}}$$

and taking expectations yields, on the left hand side,

$$p^{-1}\, \frac{E_{f_X}[|X|^p]}{E_{f_X}[|X|^p]} + q^{-1}\, \frac{E_{f_Y}[|Y|^q]}{E_{f_Y}[|Y|^q]} = p^{-1} + q^{-1} = 1$$

and on the right hand side

$$\frac{E_{f_{X,Y}}[|XY|]}{\{E_{f_X}[|X|^p]\}^{1/p}\, \{E_{f_Y}[|Y|^q]\}^{1/q}}$$

and the result follows.

Theorem (CAUCHY-SCHWARZ INEQUALITY)
Suppose that X and Y are two random variables.

$$|E_{f_{X,Y}}[XY]| \leq E_{f_{X,Y}}[|XY|] \leq \left\{E_{f_X}[|X|^2]\right\}^{1/2} \left\{E_{f_Y}[|Y|^2]\right\}^{1/2}$$

Proof Set $p = q = 2$ in the Hölder Inequality.

Corollaries:

(a) Let μ_X and μ_Y denote the expectations of X and Y respectively. Then, by the Cauchy-Schwarz inequality

$$|E_{f_{X,Y}}[(X-\mu_X)(Y-\mu_Y)]| \leq \left\{E_{f_X}[(X-\mu_X)^2]\right\}^{1/2} \left\{E_{f_Y}[(Y-\mu_Y)^2]\right\}^{1/2}$$

so that

$$E_{f_{X,Y}}[(X-\mu_X)(Y-\mu_Y)] \leq E_{f_X}[(X-\mu_X)^2] E_{f_Y}[(Y-\mu_Y)^2]$$

and hence

$$\left\{Cov_{f_{X,Y}}[X,Y]\right\}^2 \leq Var_{f_X}[X]\, Var_{f_Y}[Y].$$

(b) **Lyapunov's Inequality**: Define $Y = 1$ with probability one. Then, for $1 < p < \infty$

$$E_{f_X}[|X|] \leq \{E_{f_X}[|X|^p]\}^{1/p}.$$

Let $1 < r < p$. Then

$$E_{f_X}[|X|^r] \leq \{E_{f_X}[|X|^{pr}]\}^{1/p}$$

and letting $s = pr > r$ yields

$$E_{f_X}[|X|^r] \leq \{E_{f_X}[|X|^s]\}^{r/s}$$

so that

$$\{E_{f_X}[|X|^r]\}^{1/r} \leq \{E_{f_X}[|X|^s]\}^{1/s}$$

for $1 < r < s < \infty$.

Theorem (MINKOWSKI'S INEQUALITY)
Suppose that X and Y are two random variables, and $1 \leq p < \infty$. Then

$$\left\{E_{f_{X,Y}}[|X+Y|^p]\right\}^{1/p} \leq \{E_{f_X}[|X|^p]\}^{1/p} + \{E_{f_Y}[|Y|^p]\}^{1/p}$$

Proof Write

$$\begin{aligned} E_{f_{X,Y}}[|X+Y|^p] &= E_{f_{X,Y}}[|X+Y||X+Y|^{p-1}] \\ &\leq E_{f_{X,Y}}[|X||X+Y|^{p-1}] + E_{f_{X,Y}}[|Y||X+Y|^{p-1}] \end{aligned}$$

by the triangle inequality $|x+y| \leq |x| + |y|$. Using Hölder's Inequality on the terms on the right hand side, for q selected to satisfy $1/p + 1/q = 1$,

$$E_{f_{X,Y}}[|X+Y|^p] \leq \{E_{f_X}[|X|^p]\}^{1/p} \left\{E_{f_{X,Y}}[|X+Y|^{q(p-1)}]\right\}^{1/q} + \{E_{f_Y}[|Y|^p]\}^{1/p} \left\{E_{f_{X,Y}}[|X+Y|^{q(p-1)}]\right\}^{1/q}$$

and dividing through by $\left\{E_{f_{X,Y}}[|X+Y|^{q(p-1)}]\right\}^{1/q}$ yields

$$\frac{E_{f_{X,Y}}[|X+Y|^p]}{\left\{E_{f_{X,Y}}[|X+Y|^{q(p-1)}]\right\}^{1/q}} \leq \{E_{f_X}[|X|^p]\}^{1/p} + \{E_{f_Y}[|Y|^p]\}^{1/p}$$

and the result follows as $q(p-1) = p$, and $1 - 1/q = 1/p$.

5.3 Jensen's Inequality

Jensen's Inequality gives a lower bound on expectations of convex functions. Recall that a function $g(x)$ is **convex** if, for $0 < \lambda < 1$,

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$$

for all x and y. Alternatively, function $g(x)$ is **convex** if

$$\frac{d^2}{dt^2}\left\{g(t)\right\}_{t=x} = g^{(2)}(x) \geq 0.$$

Conversely, $g(x)$ is **concave** if $-g(x)$ is convex.

Theorem (JENSEN'S INEQUALITY)
Suppose that X is a random variable with expectation μ, and function g is convex. Then

$$E_{f_X}\left[g(X)\right] \geq g(E_{f_X}\left[X\right])$$

with equality if and only if, for every line $a + bx$ that is a tangent to g at μ

$$P[g(X) = a + bX] = 1.$$

that is, $g(x)$ is linear.

Proof Let $l(x) = a + bx$ be the equation of the tangent at $x = \mu$. Then, for each x, $g(x) \geq a + bx$ as in the figure below.

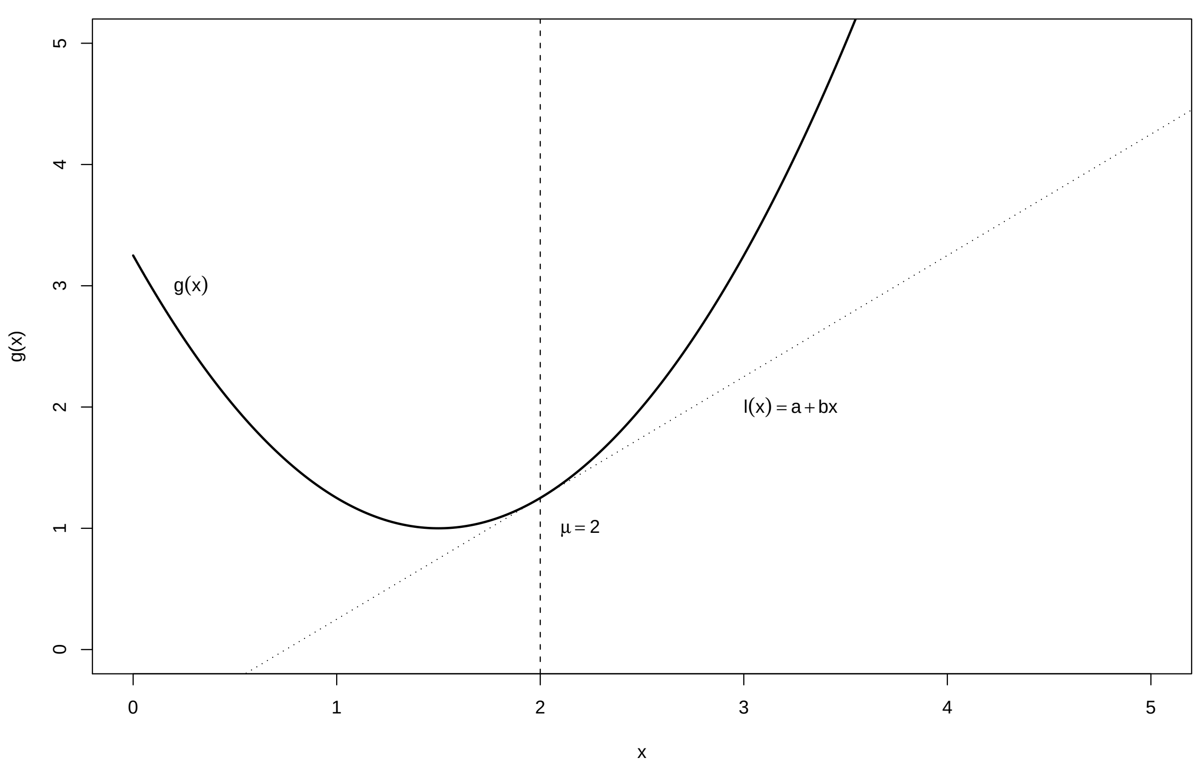

Figure 1: The function $g(x)$ and its tangent at $x = \mu$.

Thus

$$E_{f_X}[g(X)] \geq E_{f_X}[a + bX] = a + bE_{f_X}[X] = l(\mu) = g(\mu) = g(E_{f_X}[X])$$

as required. Also, **if** $g(x)$ is linear, then equality follows by properties of expectations. Suppose that

$$E_{f_X}[g(X)] = g(E_{f_X}[X]) = g(\mu)$$

but $g(x)$ is convex, but not linear. Let $l(x) = a + bx$ be the tangent to g at μ. Then by convexity

$$g(x) - l(x) > 0 \qquad \therefore \qquad \int (g(x) - l(x)) f_X(x)\, dx = \int g(x) f_X(x)\, dx - \int l(x) f_X(x)\, dx > 0$$

and hence

$$E_{f_X}[g(X)] > E_{f_X}[l(X)].$$

But $l(x)$ is linear, so $E_{f_X}[l(X)] = a + bE_{f_X}[X] = g(\mu)$, yielding the contradiction

$$E_{f_X}[g(X)] > g(E_{f_X}[X]).$$

and the result follows.

Corollary and examples:

- If $g(x)$ is **concave**, then
$$E_{f_X}[g(X)] \leq g(E_{f_X}[X])$$
- $g(x) = x^2$ is **convex**, thus
$$E_{f_X}\left[X^2\right] \geq \{E_{f_X}[X]\}^2$$
- $g(x) = \log x$ is **concave**, thus
$$E_{f_X}[\log X] \leq \log\{E_{f_X}[X]\}$$

Lemma Suppose that X is a random variable, with finite expectation μ. Let g be a non-decreasing function. Then

$$E_{f_X}[g(X)(X - \mu)] \geq 0$$

Proof By definition,

$$\begin{aligned} E_{f_X}[g(X)(X-\mu)] &= \int_{-\infty}^{\infty} g(x)(x-\mu) f_X(x)\, dx \\ &= \int_{-\infty}^{\mu} g(x)(x-\mu) f_X(x)\, dx + \int_{\mu}^{\infty} g(x)(x-\mu) f_X(x)\, dx. \end{aligned}$$

Now

$$\int_{-\infty}^{\mu} g(x)(x-\mu) f_X(x)\, dx \geq \int_{-\infty}^{\mu} g(\mu)(x-\mu) f_X(x)\, dx$$

as, on $(-\infty, \mu)$, $x < \mu$, so $x - \mu$ is **negative**, and thus as g is non-decreasing, on this range

$$g(x)(x-\mu) \geq g(\mu)(x-\mu)$$

as the left hand side is less negative than the right hand side. Similarly,

$$\int_{\mu}^{\infty} g(x)(x-\mu)f_X(x)\,dx \geq \int_{\mu}^{\infty} g(\mu)(x-\mu)f_X(x)\,dx$$

as, on (μ, ∞), $x > \mu$, so $x - \mu$ is **positive**, and thus as g is non-decreasing, on this range

$$g(x)(x-\mu) \geq g(\mu)(x-\mu)$$

as the left hand side is more positive than the right hand side. Hence

$$\begin{aligned} E_{f_X}[g(X)(X-\mu)] &\geq \int_{-\infty}^{\mu} g(\mu)(x-\mu)f_X(x)\,dx + \int_{\mu}^{\infty} g(\mu)(x-\mu)f_X(x)\,dx \\ &= g(\mu)\int_{-\infty}^{\infty}(x-\mu)f_X(x)\,dx = 0 \end{aligned}$$