Chapter 4

Inequalities

4.1 Probability Inequalities

4.1.1 Markov's inequality

Theorem 4.1.1 (Markov's inequality)**.** *If X is a non-negative random variable and $a > 0$, then the probability that X is at least a is at most the expectation of X divided by a*

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

Proof. Suppose X is a positive continuous random variable, we can write

$$\begin{aligned} E[X] &= \int_0^\infty x f_X(x) dx \\ (\text{for } a > 0) &\geq \int_a^\infty x f_X(x) dx \\ (\text{given } x > a) &\geq \int_a^\infty a f_X(x) dx \\ &= a \int_a^\infty f_X(x) dx = aP(X \geq a) \end{aligned}$$

Therefore

$$aP(X \geq a) \leq E[X] \quad \Rightarrow \quad P(X \geq a) \leq \frac{E[X]}{a}$$

Suppose X is a positive discrete random variable, depending on whether $x \geq a$, we can write

$$\begin{aligned} E[X] &= \sum_{x \geq a} xP(X = x) + \sum_{x \leq a} xP(X = x) \\ &\geq \sum_{x \geq a} xP(X = x) + 0 \\ &\geq \sum_{x \geq a} aP(X = x) + 0 \quad (\text{since } x \geq a) \\ &= a \sum_{x \geq a} P(X = x) \\ &= aP(X \geq a) \end{aligned}$$

Therefore

$$aP(X \geq a) \leq E[X] \quad \Rightarrow \quad P(X \geq a) \leq \frac{E[X]}{a}$$

□

4.1.2 Chebyshev's inequality

Theorem 4.1.2 (Chebyshev's inequality)**.** *Let X be a random variable with finite variance. Then*

$$P(|X-\mu| \geq k\sigma) = P(-k\sigma < X-\mu < k\sigma) \leq \frac{1}{k^2}$$

This is a remarkable result. It says that no matter your choice of random variable, as long as it has finite first two moments, it will not deviate from its mean by more than an explicit multiple of its standard deviation.

Proof. One way to prove Chebyshev's inequality is to apply Markov's inequality to the random variable $Y = (X-\mu)^2$ with $a = (k\sigma)^2$

$$P(|X-\mu| \geq k\sigma) = P((X-\mu)^2 \geq k^2\sigma^2) = \underbrace{P(Y \geq a) \leq \frac{E[Y]}{a}}_{\text{Markov's Inequality}} = \frac{E[(X-\mu)^2]}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

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

□

4.1.3 Hoeffding's Inequality

Theorem 4.1.3 (Hoeffding's Inequality)**.** *Suppose that $X_1, \ldots, X_n$ are independent and that, $a_i \leq X_i \leq b_i$, and $\mathbb{E}[X_i] = 0$. Then for any $\epsilon > 0$, we have the two results:*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq \epsilon\right) \leq \exp\left(-\frac{2n^2\epsilon^2}{\sum_{i=1}^{n}(b_i-a_i)^2}\right)$$

and

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2n^2\epsilon^2}{\sum_{i=1}^{n}(b_i-a_i)^2}\right)$$

One can also generalize the theorem in this way:

Let $Y_1, \ldots, Y_n$ be independent observations such that $\mathrm{E}(Y_i) = 0$ and $a_i \leq Y_i \leq b_i$. Let $\epsilon > 0$. Then, for any $t > 0$

$$P\left(\sum_{i=1}^{n} Y_i \geq \epsilon\right) \leq \exp\left(-t\epsilon + \sum_{i=1}^{n} t^2 (b_i-a_i)^2/8\right)$$

Proof. □

4.1.4 Mill's Inequality

Theorem 4.1.4 (Mill's Inequality)**.** *Let $Z \sim N(0,1)$. For any $t > 0$:*

$$\mathbb{P}(|Z| \geq t) \leq \sqrt{\frac{2}{\pi}} \frac{\exp\left(-t^2/2\right)}{t} \leq \frac{\exp\left(-t^2/2\right)}{t}$$

Proof. □

4.1.5 Chernoff bounds

Theorem 4.1.5 (Chernoff bounds)**.** *Suppose X is a random variable and we denote its moment generating function $m_X(t)$, then for any $a \in \mathbb{R}$*

$$P(X \geq a) \leq \inf_{t>0} e^{-ta} m_X(t)$$
$$P(X \leq a) \leq \inf_{t<0} e^{-ta} m_X(t)$$

Proof.

$$P(X \geq a) = P(e^{tX} \geq e^{ta}), t > 0$$
$$P(X \leq a) = P(e^{tX} \geq e^{ta}), t < 0$$

Notice that e^{tX} is a positive random variable $\forall t \in \mathbb{R}$. Therefore we can apply Markov's inequality

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}}, t > 0$$

$$P(X \leq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}}, t < 0$$

Recall that $E[e^{tX}] = m_X(t)$, hence

$$P(X \geq a) \leq \frac{m_X(t)}{e^{ta}}, t > 0$$

$$P(X \geq a) \leq \frac{m_X(t)}{e^{ta}}, t < 0$$

Taking the minimum over t and we get the result □

4.1.6 Boole's Inequality

Theorem 4.1.6 (Boole's Inequality)**.** *For a countable set of events $A_1, A_2, \cdots$, we have*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

Proof. We will prove Boole's inequality using the method of weak induction.
When $n = 1$, it is obvious that

$$P(A_1) \leq P(A_1)$$

When n, assume that

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

We will prove that this statement applies for $n + 1$. Since $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, we have

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) = P\left(\bigcup_{i=1}^{n} A_i \cup A_{n+1}\right) = P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) - P\left(\bigcup_{i=1}^{n} A_i \cap A_{n+1}\right)$$

Since by the first axiom of probability ($0 \leq P(A) \leq 1$)

$$P\left(\bigcup_{i=1}^{n} A_i \cap A_{n+1}\right) > 0$$

Then we have

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) \leq P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1})$$

With the assumption above, we could say that

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) \leq \sum_{i=1}^{n} P(A_i) + P(A_{n+1}) = \sum_{i=1}^{n+1} P(A_i)$$

□

4.1.7 Bonferroni inequalities

Boole's inequality may be generalized to find upper and lower bounds on the probability of finite unions of events. These bounds are known as Bonferroni inequalities. Assume $A_1, \cdots, A_n \in \Omega$, define

$$S_1 = \sum_{i}^{n} P(A_i)$$

$$S_2 = \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2})$$

$$\cdots$$

$$S_k = \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k})$$

Theorem 4.1.7 (Bonferroni inequalities)**.**

For odd $k \in 1, \cdots, n$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{j=1}^{k} (-1)^{j-1} S_j$$

For even $k \in 2, \cdots, n$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{j=1}^{k} (-1)^{j-1} S_j$$

4.2 Expectation Inequalities

4.2.1 Cauchy-Schwartz inequality

4.2.2 Jensen's inequality

Theorem 4.2.1 (Jensen's inequality)**.** *Suppose X is a random variable such that $a \leq X \leq b$. If $g : \mathbb{R} \to \mathbb{R}$ is convex on $[a, b]$, then*

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

if g is concave, then

$$E[g(X)] \leq g(E[X])$$

Proof. Assume X is a random variable with expectation μ, and function g is convex. Let $l(x) = a + bx$ be the equation of the tangent line at $x = \mu$. Then $\forall x, g(x) \geq a + bx$. As the figure below, where the red line is $g(x)$, the blue line is where $x = \mu$, and the green line is $l(x)$

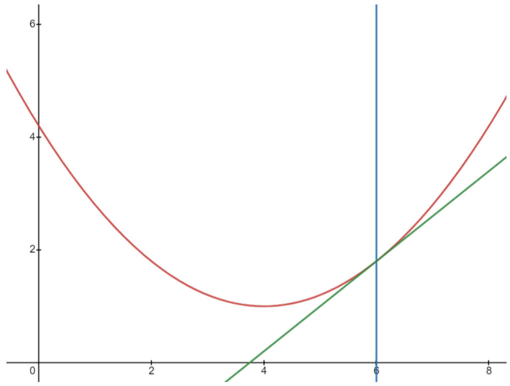

Figure 4.1

Thus

$$g(x) \geq a + bX \Rightarrow E[g(x)] \geq E[a + bX]$$
$$= a + bE[X] = l(\mu)$$

yet $l(\mu) = g(\mu)$ as in figure

$$= g(\mu) = g(E[X])$$

Now we need to prove the case where g is concave. Recall that concave function is the negative of convex function, let $h(x) = -g(x)$ be concave function

$$E[h(x)] = E[-g(x)] = -E[g(x)]$$

Since $E[g(X)] \geq g(E[X])$, $-E[g(X)] = E[h(x)] \leq g(E[X])$ $\square$

From Jensen's inequality we see that $E[X^2] \geq E[X]^2$. If X is positive, $E[1/X] \geq 1/E[X]$. Since log is concave, $E[\log(X)] \leq \log E[X]$