\

1 Basic inequalities

Theorem 1.1 (Markov's Inequality). If X is a random variable taking only non-negative values, then for any $a > 0$

$$\operatorname{Pr}[X \geq a] \leq \frac{\mathbf{E}[X]}{a} \tag{1}$$

Proof. We show this for the discrete case only, the continuous case is similar. By definition, we have

$$\mathbf{E}[X]=\sum_x x p(x)=\sum_{x<a} x p(x)+\sum_{x \geq a} x p(x) \geq \sum_{x \geq a} a p(x)=a \mathbf{Pr}[X \geq a].$$

A slightly more intuitive form of (1) is

$$\operatorname{Pr}[X \geq a \mu] \leq \frac{1}{a} \tag{2}$$

Theorem 1.2 (Chebyshev's Inequality). If X is a random variable with mean μ and variance σ^2, then for any $a>0$,

$$\operatorname{Pr}[|X-\mu| \geq a] \leq \frac{\sigma^2}{a^2} \tag{3}$$

Proof. Let $Z=(X-\mu)^2$, then $\mathbf{E}[Z]=\sigma^2$ by definition of variance. Since $|X-\mu| \geq a$ iff $Z \geq a^2$, applying Markov's inequality completes the proof.

Again, there is a more intuitive way of writing (3):

$$\operatorname{Pr}[|X-\mu| \geq a \sigma] \leq \frac{1}{a^2} \tag{4}$$

Theorem 1.4 (One-sided Chebyshev Inequality). Let X be a random variable with $\mathbf{E}[X]=\mu$ and $\operatorname{Var}[X]=\sigma^2$, then for any $a>0$,

$$\operatorname{Pr}[X \geq \mu+a] \leq \frac{\sigma^2}{\sigma^2+a^2} \tag{6}$$

$$\operatorname{Pr}[X \leq \mu-a] \leq \frac{\sigma^2}{\sigma^2+a^2} \tag{7}$$

Proof. Let $t \geq -\mu$ be a variable. Then, $Y=(X+t)^2$ has and

$$\mathbf{E}[Y]=\mathbf{E}\left[X^2\right]+2 t \mu+t^2=\sigma^2+(t+\mu)^2$$

Thus, by Markov’s inequality we get

$$\operatorname{Pr}[X \geq \mu+a] \leq \operatorname{Pr}\left[Y \geq(\mu+a+t)^2\right] \leq \frac{\sigma^2+(t+\mu)^2}{(a+t+\mu)^2}$$

The right most expression is minimized when $t=\sigma^2 / a-\mu$, in which case it becomes $\sigma^2 /\left(\sigma^2+a^2\right)$ as desired. The other inequality is proven similarly.

Theorem 1.5 (Jenssen's inequality). Let $f(x)$ be a convex function, then

$$\mathbf{E}[f(X)] \geq f(E[X]) \tag{8}$$

Proof. Taylor's theorem gives

$$f(x)=f(\mu)+f'(\mu)(x-\mu)+f''(\xi)(x-\mu)^2/2,$$

where ξ is some number between x and μ. When $f(x)$ is convex, $f''(\xi) \geq 0$, which implies

$$f(x) \geq f(\mu)+f'(\mu)(x-\mu)$$

Consequently,

$$\mathbf{E}[f(X)] \geq f(\mu)+f'(\mu)\mathbf{E}[X-\mu]=f(\mu).$$

2 Elementary Inequalities and Asymptotic Estimates

Fact 2.1. For $p \in [0,1], (1-p) \leq e^{-p}$. The inequality is good for small p.
Fact 2.2. For any $x \in [-1,1], (1+x) \leq e^{x}$. The inequality is good for small x.
The following theorem was shown by Robbins [16].
Theorem 2.3 (Stirling's approximation).

$$n!=\sqrt{2\pi n}\left(\frac{n}{e}\right)^n(1+o(1)) \tag{10}$$

3 Chernoff bounds

Theorem 3.1 (Chernoff bound). Let X be a random variable with moment generating function $M(t)=\mathbf{E}\left[e^{tX}\right]$. Then,

$$\begin{array}{lll}\operatorname{Pr}[X \geq a] \leq e^{-ta}M(t) & \text{for all} & t>0 \\ \operatorname{Pr}[X \leq a] \leq e^{-ta}M(t) & \text{for all} & t<0.\end{array}$$

Proof. The best bound can be obtained by minimizing the function on the right hand side. We show the first relation, the second is similar. When $t>0$, by Markov's inequality we get

$$\operatorname{Pr}[X \geq a]=\operatorname{Pr}\left[e^{tX} \geq e^{ta}\right] \leq \mathbf{E}\left[e^{tX}\right]e^{-ta}$$