Introduction to Probability Theory	**MSC**

Lecture 1: Inequalities

Lecturer: Internet — *Scribe: Lily Li*

Here we will cover a few good-to-know inequalities. There is an associated list of many more inequalities, but the following should be foremost in your mind at all times.

1.1 Cauchy Schwartz

The most common form of the Cauchy Schwartz inequality that you have encountered is mostly likely:

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right) \cdot \left(\sum_{i=1}^{n} b_i^2\right)$$

for sequences $a = \{a_1, ..., a_n\}$ and $b = \{b_1, ..., b_n\}$ with equality when a_i. Observe that if we consider these sequences as vectors we have the equivalent inequality: $\|a \cdot b\|^2 \leq \|a\|^2 \cdot \|b\|^2$. If you prefer a more probability theoretic formulation try something like:

$$E(X, Y) \leq E(X^2) \cdot E(Y^2).$$

Observe that if X and Y are linearly independent then we have equality. (An equivalent statement of the above is to note that $COV(X, Y) \leq 1$ where COV is the covariance).

1.2 Jensen

Jensen's inequality involves convex functions g. In particular it states:

$$E(g(x)) \geq g(E(x)).$$

The best way to see this is on a picture. See Figure 1.1. Consider the point $E(x)$ on the x-axis and draw the tangent like to $g(E(x))$. This tangent line L has the form $y = a + bx$ and is coincident to the $g(x)$ at x coordinate $E(x)$. Since g is a convex function we have that the L is below g so

$$\begin{aligned}
g(x) &\geq a + bx \\
E(g(x)) &\geq E(a + bx) \\
&= a + b \cdot E(x) && \text{by linearity of expectations} \\
&= g(E(x)) && \text{since } L \text{ is coincident with } g \text{ at E(x)}
\end{aligned}$$

1.3 Markov

The following are two very rough but useful probability theoretic inequalities. Markov's inequality states:

$$\mathsf{Pr}\left[|X| \geq a\right] \leq \frac{E(|X|)}{a}.$$

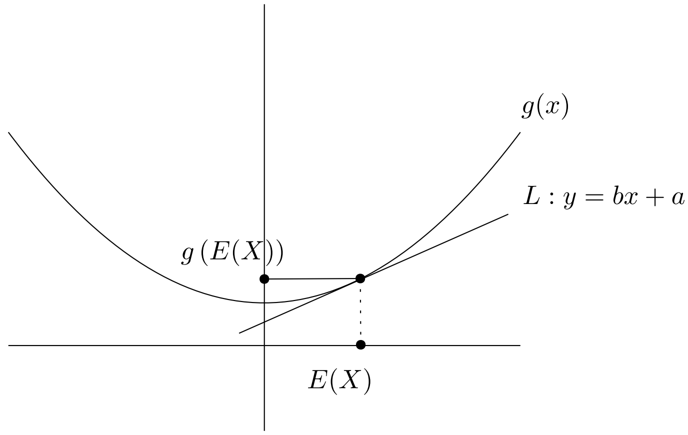

Figure 1.1: Consider the above g and the interpretation of $g(E(X))$ and $E(g(X))$.

Before we dive into the — admitted really easy — proof, you should think of the intuition for this in-equality. Imagine 100 people. At most 50, or $\frac{1}{2}$, of the people are twice the average age. Similarly, at most $\frac{1}{3}$ of the people are at most three times the average and so forth.

The proof uses the fundamental bridge between probability and expectations, and it is: $\mathsf{Pr}[|X| \geq a] = E(I_{|X| \geq a})$. This says that the probability that $|X| \geq a$ is equal to the expectation of the indicator random variable for $|X| \geq a$. Now observe that $a \cdot I_{|X| \geq a} \leq |X|$ (since indicator random variables can only be 0 or 1 consider both cases: if $I_{|X| \geq a} = 0$ then the LHS equals zero; conversely if $I_{|X| \geq a} = 1$ then $a \cdot 1 \leq |X|$). Applying the expectation function to both sides of the inequality gives you Markov's inequality.

1.4 Chebyschev

Chebyschev's inequality states:

$$\mathsf{Pr}[|X - \mu| \geq a] \leq \frac{Var(X)}{a^2}$$

and can be proved from Markov's inequality. By squaring $|X - \mu| \geq a$ we obtain $(X - \mu)^2 \geq a^2$ (this is fine since a is non-negative). Next apply Markov's inequality to obtain

$$\mathsf{Pr}\left[(X - \mu)^2 \geq a^2\right] \leq \frac{E\left((X - \mu)^2\right)}{a^2}.$$

Finally, it suffices to notice that $Var(X) = E\left((X - \mu)^2\right)$. Another more intuitive way to write this inequality is to replace a with $c \cdot \sigma$ where σ is the standard deviation. Then we have

$$\mathsf{Pr}\left[(X - \mu)^2 \geq c \cdot \sigma\right] \leq \frac{1}{c^2}.$$

You should read this as: there is less than $\frac{1}{c^2}$ of the people more than c standard deviations from the mean (if you compare this to your normal distribution, you see this inequality is indeed quite weak).