

1 Important Inequalities

Most of the applications we have seen in this course have involved the exact computations of the expected values and probabilities. In applications, the exact values are often hard to compute or simply unknown because we do not have all the required information to do an exact computation. We introduce several important inequalities that allow us to bound and approximate quantities of interest. In several cases, these inequalities are sufficient to make interesting conclusions.

1.1 Integral Inequalities

We first state several common integral inequalities we will see in any analysis course, interpreted in the language of probability theory. These inequalities are used everywhere in probability to establish bounds.

The first inequality is a bound on the expected value of the product of random variables in terms of the expected values of each component separately.

Theorem 1 (*Hölder's Inequality*)

Let X and Y be any random variables and let $p, q \geq 1$ satisfy

$$\frac{1}{p} + \frac{1}{q} = 1.$$

We have

$$|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{\frac{1}{p}} (\mathbb{E}[|Y|^q])^{\frac{1}{q}}$$

Hölder's inequality is a generalization of a well known inequality we have seen before when $p = q = 2$.

Corollary 1 (*Cauchy–Schwarz Inequality*)

For any two random variables X and Y ,

$$|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|] \leq (\mathbb{E}[X^2])^{\frac{1}{2}} (\mathbb{E}[Y^2])^{\frac{1}{2}}$$

The next identity allows us to bound the product of functions of the same random variable without appealing to higher moments.

Theorem 2 (*Covariance Inequality*)

Let X be any random variable and let $g(x)$ and $h(x)$ be any functions such that $\mathbb{E}[g(X)]$, $\mathbb{E}[h(X)]$ and $\mathbb{E}[g(X)h(X)]$ exist.

- *Negative Correlation:* If g is non-decreasing and h is non-increasing, then

$$\mathbb{E}[g(X)h(X)] \leq \mathbb{E}[g(X)] \mathbb{E}[h(X)]$$

- *Positive Correlation:* If g and h are either both non-decreasing or both non-increasing, then

$$\mathbb{E}[g(X)h(X)] \geq \mathbb{E}[g(X)] \mathbb{E}[h(X)]$$

Remark 1. The intuition is clear. In the first case, the variables $g(X)$ and $h(X)$ are negatively correlated since they move in opposite directions, so

$$\text{Cov}(g(X), h(X)) = \mathbb{E}[g(X)h(X)] - \mathbb{E}[g(X)] \mathbb{E}[h(X)] \leq 0.$$

And in the second case, they are positively correlated, so

$$\text{Cov}(g(X), h(X)) = \mathbb{E}[g(X)h(X)] - \mathbb{E}[g(X)] \mathbb{E}[h(X)] \geq 0.$$

We now provide an inequality for the sums of random variables. We can interpret it as a generalization of the triangle inequality.

Theorem 3 (*Minkowski's Inequality*)

Let X and Y be any two random variables. For $1 \leq p < \infty$,

$$(\mathbb{E}[|X + Y|^p])^{\frac{1}{p}} \leq (\mathbb{E}[|X|^p])^{\frac{1}{p}} + (\mathbb{E}[|Y|^p])^{\frac{1}{p}}.$$

Lastly, we remind ourselves of an essential integral inequality relating to the expected value of convex functions.

Theorem 4 (*Jensen's Inequality*)

For any random variables X . If $f(x)$ is a convex function, then

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]).$$

Remark 2. For concave functions, the inequality goes the other direction.

1.2 Connection Between Tail Probabilities and Moments

We saw before that the distribution of the random variable determines its expected value. The expected value does not determine the distribution, but it can be used to bound probabilities of some events. In particular, the moments of the random variable can be used to control the probability that a random variable takes extreme outlying values away from its expected value.

Theorem 5 (*Markov's Inequality*)

For any random variable and constant $a > 0$,

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

Remark 3. This is intuitive, because a small expected value means that the probability that $|X|$ is big should be small. Furthermore, the more extreme the outlier is (which is measured by a), the more rare it should be. For example, if X denotes income of a randomly selected person in a city, then

$$\mathbb{P}(X \geq 2 \mathbb{E}[X]) \leq \frac{1}{2}$$

since we can't have more than half the people in the city making twice the average income.

This is a good first step to controlling the tail events since it only requires some information about the first moment $\mathbb{E}[|X|]$. However, if we have more information such as control of the second moment, then we have better control of the tails.

Corollary 2 (*Chebyshev's Inequality*)

Let X have mean μ . Then for any constant $a > 0$,

$$\mathbb{P}(|X - \mu| \geq a) \leq \frac{\mathbb{E}[(X - \mu)^2]}{a^2} = \frac{\text{Var}(X)}{a^2}.$$

Remark 4. Notice that the tail probability in Chebyshev's Inequality is now bounded by a^{-2} instead of a^{-1} so it gives us better control than Markov's inequality for large a .

If we have even more information about the moments, then we can upgrade the bound to give us exponential control of the tails.

Corollary 3 (Chernoff Bound)

For any random variable and constants $t > 0$ and $a > 0$

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}} = M_X(t)e^{-ta}.$$

There is an extra parameter t in the statement of the Chernoff bound, so we can minimize the upper bound as a function of t to get the better control of the tail probabilities.

1.3 Example Problems

Problem 1.1. An insurance company sells 1,000 independent health insurance policies. Suppose that for any policy holder, the expected claim is 5,000 but the variance is 2,000,000 since the costs of the treatment can vary a lot. What is a bound on the probability that the insurance company has to pay more than 10,000,000 in a given year.

Solution 1.1. Notice that $T = \sum_{i=1}^{1000} X_i$ the total claim amount satisfies

$$\mathbb{E}(T) = \sum_{i=1}^{1000} \mathbb{E}[X_i] = 5,000,000 \text{ and } \text{Var}(T) = \sum_{i=1}^{1000} \text{Var}\left(X_i\right) = 2,000,000,000.$$

Therefore, by Chebyshev's inequality

$$\begin{aligned} \mathbb{P}(T > 10,000,000) &= \mathbb{P}(T - 5,000,000 > 5,000,000) \\ &\leq \mathbb{P}(|T - 5,000,000| > 5,000,000) \\ &\leq \frac{\text{Var}(T)}{5,000,000^2} = \frac{2,000,000,000}{5,000,000^2} = 0.00008. \end{aligned}$$

Remark 5. If we used Markov's inequality instead, we would have gotten

$$\mathbb{P}(T > 10,000,000) \leq \frac{\mathbb{E}(T)}{10,000,000} = \frac{1}{2}$$

which is true, but not as good of a bound as if we used information about the variance.

Problem 1.2. Suppose a machine produces metal rods whose lengths (in cm) are non-negative random variables L with mean 100.

1. Use Markov's inequality to bound the probability that a rod exceeds 150 cm.
2. Suppose that we now know that the variance of the rods is 25 cm^2 . Use Chebyshev's inequality to bound the probability that a rod exceeds 150 cm.

Solution 1.2.

Part 1: By Markov's inequality (since the lengths of the rods are non-negative)

$$\mathbb{P}(L > 150) \leq \frac{\mathbb{E}[L]}{150} = \frac{2}{3}.$$

Part 2: Since $\{|L - 100| > 50\} = \{L > 150\} \cup \{L < 50\}$, by Chebyshev's inequality,

$$\mathbb{P}(L > 150) \leq \mathbb{P}(|L - 100| > 50) \leq \frac{\text{Var}(L)}{50^2} = \frac{1}{100}.$$

Problem 1.3. Let Y be the average income (in thousands) of 100 independently surveyed people from a population with mean 50 and variance 100. Use Chebyshev's inequality to bound

$$\mathbb{P}(|Y - 50| \geq 10).$$

Solution 1.3. We have that Y is the average income of 100 people, so

$$\text{Var}(Y) = \text{Var}\left(\frac{1}{100} \sum_{i=1}^{100} X_i\right) = \frac{100 \text{Var}(X)}{100^2} = 1$$

Therefore,

$$\mathbb{P}(|Y - 50| \geq 10) \leq \frac{\text{Var}(Y)}{10^2} = \frac{1}{100},$$

which is quite small even though the standard deviation of a single person is 10.

1.4 Proofs of Key Results

Problem 1.4. Prove Young's inequality: Suppose that $p, q > 1$

$$\frac{1}{p} + \frac{1}{q} = 1.$$

For any positive numbers a and b ,

$$\frac{1}{p}a^p + \frac{1}{q}b^q \geq ab. \tag{1}$$

Equality is attained when $a^p = b^q$.

Solution 1.4. This is a calculus proof. For any fixed $b > 0$, consider the function

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

To prove Young's inequality, it suffices to show that $g(a) \geq 0$. Since $p \geq 1$, the function $g(a)$ is convex in a , so the global minimum of $g(a)$ is attained at

$$g'(a) = 0 \implies a^{p-1} - b = 0 \implies b = a^{p-1} \stackrel{p=q(p-1)}{\implies} a^p = b^q.$$

Using the fact

$$\frac{1}{p} + \frac{1}{q} = 1 \iff (p-1)q = p$$

substituting the critical point condition $b = a^{p-1}$ in implies that

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

is attained at 0. At the boundary $a = 0$ or $b = 0$, Young's inequality holds trivially. Furthermore, if using the fact that $(p-1)q = p$, at the critical point $b = a^{p-1} \implies a^p = b^q$ we have equality.

Problem 1.5. Prove Theorem 1 (Hölder's Inequality).

Solution 1.5. The first inequality in Hölder's is trivial and follows from the fact that $-|XY| \leq XY \leq |XY|$, so $|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|]$ by the linearity of expectation.

If either $\mathbb{E}[|X|^p] = 0$ or $\mathbb{E}[|Y|^q] = 0$, then the statement is trivial. If they are both positive, we define

$$a = \frac{|X|}{(\mathbb{E}[|X|^p])^{\frac{1}{p}}} \quad \text{and} \quad b = \frac{|Y|}{(\mathbb{E}[|Y|^q])^{\frac{1}{q}}}$$

so that Young's inequality (1) implies that

$$\frac{1}{p} \frac{|X|^p}{\mathbb{E}[|X|^p]} + \frac{1}{q} \frac{|Y|^q}{\mathbb{E}[|Y|^q]} \geq \frac{|X||Y|}{(\mathbb{E}[|X|^p])^{\frac{1}{p}} (\mathbb{E}[|Y|^q])^{\frac{1}{q}}}.$$

Taking expectation of both sides and using the fact that $\frac{1}{p} + \frac{1}{q} = 1$ implies the left hand side is 1, so

$$1 \geq \frac{\mathbb{E}[|X||Y|]}{(\mathbb{E}[|X|^p])^{\frac{1}{p}} (\mathbb{E}[|Y|^q])^{\frac{1}{q}}} = \frac{\mathbb{E}[|XY|]}{(\mathbb{E}[|X|^p])^{\frac{1}{p}} (\mathbb{E}[|Y|^q])^{\frac{1}{q}}}$$

and Hölder's inequality follows from a simple rearrangement.

Problem 1.6. Prove Theorem 2 (Covariance Inequality).

Solution 1.6. Let Y be an independent copy of X . We will show that

$$\mathbb{E}[g(X)h(X)] - \mathbb{E}[g(X)]\mathbb{E}[h(X)] = \text{Cov}(g(X), h(X)) = \frac{1}{2} \mathbb{E}[(g(X) - g(Y))(h(X) - h(Y))]. \quad (2)$$

To see how this implies the covariance inequality, notice that if g is non-decreasing and h is non-increasing, then we have for every x, y

$$(g(x) - g(y))(h(x) - h(y)) \leq 0$$

and the opposite inequality holds if g and h are either both non-decreasing or both non-increasing. Taking expectation and applying (2) finishes the proof.

To show (2), we have

$$\begin{aligned} & \frac{1}{2} \mathbb{E}[(g(X) - g(Y))(h(X) - h(Y))] \\ &= \frac{1}{2} \left[\mathbb{E}[g(X)h(X)] - \mathbb{E}[g(X)h(Y)] - \mathbb{E}[g(Y)h(X)] + \mathbb{E}[g(Y)h(Y)] \right] \\ &= \frac{1}{2} \left[\mathbb{E}[g(X)h(X)] - \mathbb{E}[g(X)]\mathbb{E}[h(Y)] - \mathbb{E}[g(Y)]\mathbb{E}[h(X)] + \mathbb{E}[g(Y)h(Y)] \right] \\ &= \frac{1}{2} \left[\mathbb{E}[g(X)h(X)] - \mathbb{E}[g(X)]\mathbb{E}[h(X)] - \mathbb{E}[g(X)]\mathbb{E}[h(X)] + \mathbb{E}[g(X)h(X)] \right] \\ &= \mathbb{E}[g(X)h(X)] - \mathbb{E}[g(X)]\mathbb{E}[h(X)] = \text{Cov}(g(X), h(X)). \end{aligned}$$

where we used that X and Y are independent in the second equality and have the same distribution in the third equality.

Problem 1.7. Prove Theorem 3 (Minkowski's Inequality).

Solution 1.7. The case when $p = 1$ is simply the triangle inequality. Let $p > 1$. This is an application of Hölder's inequality. We have for $p \geq 1$,

$$\mathbb{E}[|X + Y|^p] = \mathbb{E}[|X + Y||X + Y|^{p-1}] \leq \mathbb{E}[|X||X + Y|^{p-1}] + \mathbb{E}[|Y||X + Y|^{p-1}]$$

by the triangle inequality $|X + Y| \leq |X| + |Y|$. We define $q = \frac{p}{p-1}$ so that $\frac{1}{p} + \frac{1}{q} = 1$. Applying Hölder's inequality to each term on the right hand side implies that

$$\mathbb{E}[|X + Y|^p] \leq \mathbb{E}[|X|^p]^{\frac{1}{p}} \mathbb{E}[|X + Y|^{q(p-1)}]^{\frac{1}{q}} + \mathbb{E}[|Y|^p]^{\frac{1}{p}} \mathbb{E}[|X + Y|^{q(p-1)}]^{\frac{1}{q}}.$$

Using the fact that $q(p-1) = p$ and $1 - \frac{1}{q} = \frac{1}{p}$ and dividing both sides by $\mathbb{E}[|X + Y|^{q(p-1)}]^{\frac{1}{q}} = \mathbb{E}[|X + Y|^p]^{\frac{1}{q}}$ and implies that

$$(\mathbb{E}[|X + Y|^p])^{1 - \frac{1}{q}} = (\mathbb{E}[|X + Y|^p])^{\frac{1}{p}} \leq (\mathbb{E}[|X|^p])^{\frac{1}{p}} + (\mathbb{E}[|Y|^p])^{\frac{1}{p}}.$$

Remark 6. Notice that by Jensen's inequality applied to the sum,

$$\mathbb{E}[|X + Y|^p] = 2^p \mathbb{E}\left[\left|\frac{X + Y}{2}\right|^p\right] \leq 2^{p-1}(\mathbb{E}|X|^p + \mathbb{E}|Y|^p)$$

so the left hand side is finite whenever the right hand side is. Furthermore, if $\mathbb{E}[|X + Y|^p]$ is infinite or zero, the statement is trivial so we can exclude the case we divide by 0 or ∞ .

Problem 1.8. Prove Theorem 4 (Jensen's Inequality).

Solution 1.8. We provide a proof when f is differentiable. The statement holds even if f is not differentiable, but it involves using subdifferential lines instead of the tangent line. If $f(x)$ is convex and differentiable, then its tangent line at any point y is below f ,

$$f(x) \geq f(y) + f'(y)(x - y).$$

This holds for all x and y , so it holds for the random choice $x = X$ and $y = \mathbb{E}[X]$

$$f(X) - f(\mathbb{E}[X]) \geq f'(\mathbb{E}[X])(X - \mathbb{E}[X]).$$

Taking expectations gives

$$\mathbb{E}[f(X) - f(\mathbb{E}[X])] \geq \mathbb{E}[f'(\mathbb{E}[X])(X - \mathbb{E}[X])].$$

The right hand side is equal to zero since

$$\mathbb{E}[f'(\mathbb{E}[X])(X - \mathbb{E}[X])] = f'(\mathbb{E}[X]) \mathbb{E}[(X - \mathbb{E}[X])] = f'(\mathbb{E}[X])(\mathbb{E}[X] - \mathbb{E}[X]) = 0.$$

Therefore, by linearity

$$\mathbb{E}[f(X) - f(\mathbb{E}[X])] \geq 0 \implies \mathbb{E}[f(X)] \geq \mathbb{E}[f(\mathbb{E}[X])] = f(\mathbb{E}[X]).$$

Problem 1.9. Prove Theorem 5, Markov's inequality.

Solution 1.9. For any $a > 0$, we define $Y = \frac{|X|}{a}$. It suffices to show that $\mathbb{P}(Y \geq 1) \leq \mathbb{E}[Y]$ since

$$\mathbb{P}(|X| \geq a) = \mathbb{P}(Y \geq 1) \leq \mathbb{E}[Y] = \frac{\mathbb{E}[|X|]}{a}.$$

Notice that if $Y \geq 0$, then we always have

$$\mathbb{1}(Y \geq 1) = \begin{cases} 1 & Y \geq 1 \\ 0 & Y < 1 \end{cases} \leq Y.$$

Monotonicity of the expected value implies that

$$\mathbb{E}[\mathbb{1}(Y \geq 1)] \leq \mathbb{E}[Y] \implies \mathbb{P}(Y \geq 1) \leq \mathbb{E}[Y].$$

Problem 1.10. Prove Corollary 2 and Corollary 3. That is, for any $t > 0$ and $a > 0$, show that

1.

$$\mathbb{P}(|X - \mu| \geq a) \leq \frac{\text{Var}(X)}{a^2}$$

2.

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}}$$

Solution 1.10. Both of these statements are direct consequences of Markov's inequality.

Part 1: We have

$$\mathbb{P}(|X - \mu| \geq a) = \mathbb{P}(|X - \mu|^2 \geq a^2),$$

so Markov's inequality implies that

$$\mathbb{P}(|X - \mu|^2 \geq a^2) \leq \frac{\mathbb{E}[|X - \mu|^2]}{a^2} = \frac{\text{Var}(X)}{a^2}.$$

Part 2: We have

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq e^{ta}),$$

so Markov's inequality implies that

$$\mathbb{P}(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}} = M_X(t)e^{-ta}.$$