CHAPTER 15

Probability inequalities

We already used several types of inequalities, and in this Chapter we give a more systematic description of the inequalities and bounds used in probability and statistics.

15.1.* Boole's inequality, Bonferroni inequalities

Boole's inequality(or the *union bound*) states that for any at most countable collection of events, the probability that at least one of the events happens is no greater than the sum of the probabilities of the events in the collection.

Proposition 15.1 (Boole's inequality)

Suppose $(S, \mathcal{F}, \mathbb{P})$ is a probability space, and $E_1, E_2, \ldots \in \mathcal{F}$ are events. Then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) \leqslant \sum_{i=1}^{\infty} \mathbb{P}\left(E_i\right).$$

Proof. We only give a proof for a finite collection of events, and we mathematical induction on the number of events.

For the $n=1$ we see that

$$\mathbb{P}\left(E_1\right) \leqslant \mathbb{P}\left(E_1\right).$$

Suppose that for some n and any collection of events $E_1, \ldots, E_n$ we have

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leqslant \sum_{i=1}^{n} \mathbb{P}\left(E_i\right).$$

Recall that by (2.1.1) for any events A and B we have

$$\mathbb{P}(A \cup B)=\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A \cap B).$$

We apply it to $A=\bigcup_{i=1}^{n} E_i$ and $B=E_{n+1}$ and using the associativity of the union $\bigcup_{i=1}^{n+1} E_i = A \cup B$, we get that

$$\mathbb{P}\left(\bigcup_{i=1}^{n+1} E_i\right)=\mathbb{P}\left(\bigcup_{i=1}^{n} E_i)+\mathbb{P}(E_{n+1}\right)-\mathbb{P}\left(\left(\bigcup_{i=1}^{n} E_i\right) \bigcap E_{n+1}\right).$$

By the first axiom of probability

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

and therefore we have

$$\mathbb{P}\left(\bigcup_{i=1}^{n+1} E_i\right) \leqslant \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) + \mathbb{P}\left(E_{n+1}\right).$$

Thus using the *induction hypothesis* we see that

$$\mathbb{P}\left(\bigcup_{i=1}^{n+1} E_i\right) \leqslant \sum_{i=1}^{n} \mathbb{P}\left(E_i\right) + \mathbb{P}\left(E_{n+1}\right) = \sum_{i=1}^{n+1} \mathbb{P}\left(E_i\right).$$

□

One of the interpretations of Boole's inequality is what is known as σ-*sub-additivity* in measure theory applied here to the probability measure $\mathbb{P}$.

Boole's inequality can be extended to get lower and upper bounds on probability of unions of events known as *Bonferroni inequalities*. As before suppose $(S, \mathcal{F}, \mathbb{P})$ is a probability space, and $E_1, E_2, ...E_n \in \mathcal{F}$ are events. Define

$$\begin{aligned}
S_1 &:= \sum_{i=1}^{n} \mathbb{P}\left(E_i\right), \\
S_2 &:= \sum_{1 \leqslant i < j \leqslant n} \mathbb{P}\left(E_i \cap E_j\right) \\
S_k &:= \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \mathbb{P}\left(E_{i_1} \cap \cdots \cap E_{i_k}\right), k = 3, ..., n.
\end{aligned}$$

Proposition 15.2 (Bonferroni inequalities)

For odd k in $1, ..., n$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leqslant \sum_{j=1}^{k} (-1)^{j-1} S_j,$$

for even k in $2, ..., n$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \geqslant \sum_{j=1}^{k} (-1)^{j-1} S_j.$$

We omit the proof which starts with considering the case $k = 1$ for which we need to show

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leqslant \sum_{j=1}^{1}(-1)^{j-1}S_j = S_1 = \sum_{i=1}^{n}\mathbb{P}\left(E_i\right),$$

which is Boole's inequality. When $k = 2$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \geqslant \sum_{j=1}^{2}(-1)^{j-1}S_j = S_1 - S_1 = \sum_{i=1}^{n}\mathbb{P}\left(E_i\right) - \sum_{1\leqslant i<j\leqslant n}\mathbb{P}\left(E_i \cap E_j\right).$$

which for $n = 2$ is the inclusion-exclusion identity (Proposition 2.2).

Example 15.1. Suppose we place n distinguishable balls into m distinguishable boxes at random $(n > m)$. Let E be the event that a box is empty. The sample space can be described as

$$\Omega = \{\omega = (\omega_1, ..., \omega_n) : 1 \leqslant \omega_i \leqslant m\}$$

with $\mathbb{P}(\omega) = \frac{1}{m^n}$. Denote $E_l := \{\omega : \omega_i \neq l \text{ for all } i = 1, ..., n\}$ for $l = 1, 2, ..., m$. Then, $E = E_1 \cup ... \cup E_{m-1}$ since E_m is empty, and we can include it or not, this does not change the result.

We can see that for any m we have

$$\mathbb{P}\left(E_{i_1} \cup ... \cup E_{i_k}\right) = \frac{(m-k)^n}{k^n} = \left(1 - \frac{k}{m}\right)^n.$$

Then we can use the inclusion-exclusion principle to get

$$\mathbb{P}(E) = m\left(1 - \frac{1}{m}\right)^n - \binom{m}{2}\left(1 - \frac{2}{m}\right)^n + ... + (-1)^{m-2}\binom{m}{m-1}\left(1 - \frac{m-1}{m}\right)^n$$

The last term is zero, since all boxes can not be empty. The expression is quite complicated. But if we use Bonferroni inequalities we see that

$$m\left(1 - \frac{1}{m}\right)^n - \binom{m}{2}\left(1 - \frac{2}{m}\right)^n \leqslant \mathbb{P}(E) \leqslant m\left(1 - \frac{1}{m}\right)^n$$

This gives a good estimate when n is large compared to m. For example, if $m = 10$ then

$$10 \cdot (0.9)^n - 45 \cdot (0.8)^n \leqslant \mathbb{P}(E) \leqslant 10 \cdot (0.9)^n.$$

In particular, for $n = 50$, then $45 \cdot (0.8)^{50} = 0.00064226146$, which is the difference between the left and right sides of the estimates. This gives a rather good estimate.

15.2. Markov's inequality

Proposition 15.3 (Markov's inequality)

Suppose X is a nonnegative random variable, then for any $a > 0$ we have

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}X}{a}$$

PROOF. We only give the proof for a continuous random variable, the case of a discrete random variable is similar. Suppose X is a positive continuous random variable, we can write

$$\begin{aligned}\mathbb{E}X &= \int_{-\infty}^{\infty} x f_X(x)\,dx \overset{X\geqslant 0}{=} \int_0^{\infty} x f_X(x)\,dx \\ &\overset{a>0}{\geqslant} \int_a^{\infty} x f_X(x)\,dx \overset{x>a}{\geqslant} \int_a^{\infty} a f_X(x)\,dx = a\int_a^{\infty} f_X(x)\,dx = a\mathbb{P}(X \geqslant a).\end{aligned}$$

Therefore

$$a\mathbb{P}(X \geqslant a) \leqslant \mathbb{E}X$$

which is what we wanted to prove. □

Example 15.2. First we observe that Boole's inequality can be interpreted as expectations of the number of occurred events. Suppose $(S, \mathcal{F}, \mathbb{P})$ is a probability space, and $E_1, E_2, ... \in \mathcal{F}$ are events. Define

$$X_i := \begin{cases} 1, & \text{if } E_i \text{ occurs} \\ 0, & \text{otherwise.} \end{cases}$$

Then $X := X_1 + ... + X_n$ is the number of events that occur. Then

$$\mathbb{E}X = \mathbb{P}(E_1) + ... + \mathbb{P}(E_n).$$

Now we would like to prove Boole's inequality using Markov's inequality. Note that X is a nonnegative random variable, so we can apply Markov's inequality. For $a = 1$ we get

$$\mathbb{P}(X \geqslant 1) \leqslant \mathbb{E}X = \mathbb{P}(E_1) + ... + \mathbb{P}(E_n).$$

Finally we see that the event $X \geqslant 1$ means that at least one of the events $E_1, E_2, ...E_n$ occur, so

$$\mathbb{P}(X \geqslant 1) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right),$$

therefore

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \mathbb{P}(X \geqslant 1) \leqslant \mathbb{E}X = \mathbb{P}(E_1) + ... + \mathbb{P}(E_n)$$

which completes the proof.

Example 15.3. Suppose $X \sim \operatorname{Binom}(n, p)$. We would like to use Markov's inequality to find an upper bound on $\mathbb{P}(X \geqslant qn)$ for $p < q < 1$.

Note that X is a nonnegative random variable and $\mathbb{E}X = np$. By Markov's inequality, we have

$$\mathbb{P}(X \geqslant qn) \leqslant \frac{\mathbb{E}X}{qn} = \frac{p}{q}.$$

15.3. Chebyshev's inequality

Here we revisit Chebyshev's inequality Proposition 14.1 we used previously. This results shows that the difference between a random variable and its expectation is controlled by its variance. Informally we can say that it shows how far the random variable is from its mean on average.

Proposition 15.4 (Chebyshev's inequality)

Suppose X is a random variable, then for any $b > 0$ we have

$$\mathbb{P}\left(|X - \mathbb{E}X| \geqslant b\right) \leqslant \frac{\operatorname{Var}(X)}{b^2}.$$

PROOF. Define $Y := (X - \mathbb{E}X)^2$, then Y is a nonnegative random variable and we can apply Markov's inequality (Proposition 15.3) to Y. Then for $b > 0$ we have

$$\mathbb{P}\left(Y \geqslant b^2\right) \leqslant \frac{\mathbb{E}Y}{b^2}.$$

Note that

$$\begin{aligned} &\mathbb{E}Y = \mathbb{E}\left(X - \mathbb{E}X\right)^2 = \operatorname{Var}(X), \\ &\mathbb{P}\left(Y \geqslant b^2\right) = \mathbb{P}\left((X - \mathbb{E}X)^2 \geqslant b^2\right) = \mathbb{P}\left(|X - \mathbb{E}X| \geqslant b\right) \end{aligned}$$

which completes the proof. □

Example 15.4. Consider again $X \sim \operatorname{Binom}(n, p)$. We now will use Chebyshev's inequality to find an upper bound on $\mathbb{P}(X \geqslant qn)$ for $p < q < 1$.

Recall that $\mathbb{E}X = np$. By Chebyshev's inequality with $b = (q - p)\, n > 0$ we have

$$\begin{aligned} &\mathbb{P}(X \geqslant qn) = \mathbb{P}(X - np \geqslant (q-p)\, n) \leqslant \mathbb{P}(|X - np| \geqslant (q-p)\, n) \\ &\leqslant \frac{\operatorname{Var}(X)}{((q-p)\, n)^2} = \frac{p(1-p)\, n}{((q-p)\, n)^2} = \frac{p(1-p)}{(q-p)^2\, n}. \end{aligned}$$

15.4. Chernoff bounds

Proposition 15.5 (Chernoff bounds)

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

$$\mathbb{P}(X \geqslant a) \leqslant \min_{t>0} e^{-ta} m_X(t),$$
$$\mathbb{P}(X \leqslant a) \leqslant \min_{t<0} e^{-ta} m_X(t).$$

Proof.

$$\mathbb{P}(X \geqslant a) = \mathbb{P}\left(e^{tX} \geqslant e^{ta}\right), t > 0,$$
$$\mathbb{P}(X \leqslant a) = \mathbb{P}\left(e^{tX} \geqslant e^{ta}\right), t < 0.$$

Note that note that e^{tX} is a positive random variable for any $t \in \mathbb{R}$. Therefore we can apply Markov's inequality (Proposition 15.3) to see that

$$\mathbb{P}(X \geqslant a) = \mathbb{P}\left(e^{tX} \geqslant e^{ta}\right) \leqslant \frac{\mathbb{E}e^{tX}}{e^{ta}}, t > 0,$$
$$\mathbb{P}(X \leqslant a) = \mathbb{P}\left(e^{tX} \geqslant e^{ta}\right) \leqslant \frac{\mathbb{E}e^{tX}}{e^{ta}}, t < 0.$$

Recall that $\mathbb{E}e^{tX}$ is the moment generating function $m_X(t)$, and so we have

$$\mathbb{P}(X \geqslant a) \leqslant \frac{m_X(t)}{e^{ta}}, t > 0,$$
$$\mathbb{P}(X \leqslant a) \leqslant \frac{m_X(t)}{e^{ta}}, t < 0.$$

Taking the minimum over appropriate t we get the result.

$\square$

Example 15.5. Consider again $X \sim \operatorname{Binom}(n, p)$. We now will use Chernoff bounds for $\mathbb{P}(X \geqslant qn)$ for $p < q < 1$. Recall that in Example 13.2 we found the moment generating function for X as follows

$$m_X(t) = (pe^t + (1-p))^n.$$

Thus a Chernoff bound gives

$$\mathbb{P}(X \geqslant qn) \leqslant \min_{t>0} e^{-tqn} \left(pe^t + (1-p)\right)^n.$$

To find the minimum of $g(t) = e^{-tqn}\left(pe^t + (1-p)\right)^n$ we can take its derivative and using the only critical point of this function , we can see that the minimum on $(0, \infty)$ is achieved at t_* such that

$$e^{t_*} = \frac{q(1-p)}{(1-q)p},$$

and so

$$\begin{aligned}
g(t_*) &= \left(\frac{q(1-p)}{(1-q)p}\right)^{-qn} \left(p\frac{q(1-p)}{(1-q)p} + (1-p)\right)^{n} \\
&= \left(\frac{q(1-p)}{(1-q)p}\right)^{-qn} \left(\frac{1-p}{1-q}\right)^{n} = \left(\frac{p}{q}\right)^{qn} \left(\frac{1-p}{1-q}\right)^{-qn} \left(\frac{1-p}{1-q}\right)^{n} \\
&= \left(\frac{p}{q}\right)^{qn} \left(\frac{1-p}{1-q}\right)^{(1-q)n}.
\end{aligned}$$

Thus the Chernoff bound gives

$$\mathbb{P}(X \geqslant qn) \leqslant \left(\frac{p}{q}\right)^{qn} \left(\frac{1-p}{1-q}\right)^{(1-q)n}.$$

Example 15.6 (Comparison of Markov's, Chebyshev's inequalities and Chernoff bounds)**.** These three inequalities for the binomial random variable $X \sim \operatorname{Binom}(n,p)$ give

$$\begin{aligned}
&\text{Markov's inequality} && \mathbb{P}(X \geqslant qn) \leqslant \frac{p}{q}, \\
&\text{Chebyshev's inequality} && \mathbb{P}(X \geqslant qn) \leqslant \frac{p(1-p)}{(q-p)^2 n}, \\
&\text{Chernoff bound} && \mathbb{P}(X \geqslant qn) \leqslant \left(\frac{p}{q}\right)^{qn} \left(\frac{1-p}{1-q}\right)^{(1-q)n}.
\end{aligned}$$

Clearly the right-hand sides are very different: Markov's inequality gives a bound independent of n, and the Chernoff bound is the strongest with exponential convergence to 0 as $n \to \infty$. For example, for $p = 1/2$ and $q = 3/4$ we have

$$\begin{aligned}
&\text{Markov's inequality} && \mathbb{P}\left(X \geqslant \frac{3n}{4}\right) \leqslant \frac{2}{3}, \\
&\text{Chebyshev's inequality} && \mathbb{P}\left(X \geqslant \frac{3n}{4}\right) \leqslant \frac{4}{n}, \\
&\text{Chernoff bound} && \mathbb{P}\left(X \geqslant \frac{3n}{4}\right) \leqslant \left(\frac{16}{27}\right)^{n/4}.
\end{aligned}$$

For example, for $p = 1/3$ and $q = 2/3$ we have

$$\begin{aligned}
&\text{Markov's inequality} && \mathbb{P}\left(X \geqslant \frac{3n}{4}\right) \leqslant \frac{1}{2}, \\
&\text{Chebyshev's inequality} && \mathbb{P}\left(X \geqslant \frac{3n}{4}\right) \leqslant \frac{2}{n}, \\
&\text{Chernoff bound} && \mathbb{P}\left(X \geqslant \frac{3n}{4}\right) \leqslant 2^{-n/2}.
\end{aligned}$$

15.5. Cauchy-Schwarz inequality

This inequality appears in a number of areas of mathematics including linear algebra. We will apply it to give a different proof for the bound for correlation coefficients. Note that the Cauchy-Schwarz inequality can be easily generalized to random vectors X and Y.

Proposition 15.6 (Cauchy-Schwarz inequality)

Suppose X and Y are two random variables, then

$$(\mathbb{E}XY)^2 \leqslant \mathbb{E}X^2 \cdot \mathbb{E}Y^2,$$

and the equality holds if and only if $X = aY$ for some constant $a \in \mathbb{R}$.

Proof. Define the random variable $U := (X - sY)^2$ which is a nonnegative random variable for any $s \in \mathbb{R}$. Then

$$0 \leqslant \mathbb{E}U = \mathbb{E}(X - sY)^2 = \mathbb{E}X^2 - 2s\mathbb{E}XY + s^2\mathbb{E}Y^2.$$

Define $g(s) := \mathbb{E}X^2 - 2s\mathbb{E}XY + s^2\mathbb{E}Y^2$ which is a quadratic polynomial in s. What we know is that $g(s)$ is nonnegative for all s. Completing the square we see that

$$g(s) = \mathbb{E}Y^2 s^2 - 2\mathbb{E}XYs + \mathbb{E}X^2 = \left(\sqrt{\mathbb{E}Y^2}s - \frac{\mathbb{E}XY}{\sqrt{\mathbb{E}Y^2}}\right)^2 + \mathbb{E}X^2 - \frac{(\mathbb{E}XY)^2}{\mathbb{E}Y^2},$$

so $g(s) \geqslant 0$ for all s if and only if

$$\mathbb{E}X^2 - \frac{(\mathbb{E}XY)^2}{\mathbb{E}Y^2} \geqslant 0,$$

which is what we needed to show.

To deal with the last claim, observe that if $U > 0$ with probability one, then $g(s) = \mathbb{E}U > 0$. This happens only if

$$\mathbb{E}X^2 - \frac{(\mathbb{E}XY)^2}{\mathbb{E}Y^2} > 0.$$

And if $\mathbb{E}X^2 - \frac{(\mathbb{E}XY)^2}{\mathbb{E}Y^2} = 0$, then $g\left(\frac{\mathbb{E}XY}{\mathbb{E}Y^2}\right) = \mathbb{E}U = 0$, which only can be true if

$$X - \frac{\mathbb{E}XY}{\mathbb{E}Y^2}Y = 0,$$

that is, X is a scalar multiple of Y. □

Example 15.7. We can use the Cauchy-Schwartz inequality to prove one of the properties of correlation coefficient in Proposition 12.3.2. Namely, suppose X and Y are random variables, then $|\rho(X,Y)| \leqslant 1$. Moreover, $|\rho(X,Y)| = 1$ if and only if there are constants $a, b \in \mathbb{R}$ such that $X = a + bY$.

We will use normalized random variables as before, namely,

$$U := \frac{X - \mathbb{E}X}{\sqrt{\operatorname{Var} X}},$$
$$V := \frac{Y - \mathbb{E}Y}{\sqrt{\operatorname{Var} Y}}.$$

Then $\mathbb{E}U = \mathbb{E}V = 0, \mathbb{E}U^2 = \mathbb{E}V^2 = 1$. We can use the Cauchy-Schwartz inequality for U and V to see that

$$|\mathbb{E}UV| \leqslant \sqrt{\mathbb{E}U^2 \cdot \mathbb{E}V^2} = 1$$

and the identity holds if and only if $U = aV$ for some $a \in \mathbb{R}$.

Recall Equation (12.3.1)

$$\rho(X,Y) = \mathbb{E}(UV),$$

which gives the bound we need. Note that if $U = aV$, then

$$\frac{X - \mathbb{E}X}{\sqrt{\operatorname{Var} X}} = a\left(\frac{Y - \mathbb{E}Y}{\sqrt{\operatorname{Var} Y}}\right),$$

therefore

$$X = a\sqrt{\operatorname{Var} X}\left(\frac{Y - \mathbb{E}Y}{\sqrt{\operatorname{Var} Y}}\right) + \mathbb{E}X = a\frac{\sqrt{\operatorname{Var} X}}{\sqrt{\operatorname{Var} Y}}Y - a\frac{\sqrt{\operatorname{Var} X}}{\sqrt{\operatorname{Var} Y}}\mathbb{E}Y + \mathbb{E}X,$$

which completes the proof.

15.6. Jensen's inequality

Recall that a function $g : \mathbb{R} \longrightarrow \mathbb{R}$ is *convex* on $[a,b]$ if for each $x, y \in [a,b]$ and each $\lambda \in [0,1]$ we have

$$g(\lambda x + (1-\lambda)y) \leqslant \lambda g(x) + (1-\lambda) g(y).$$

Note that for a convex function g this property holds for any *convex linear combination* of points in $[a,b]$, that is,

$$g(\lambda_1 x_1 + ... + \lambda_n x_n) \leqslant \lambda_1 g(x_1) + ... + g(\lambda_n x_n),$$
$$\lambda_1 + ... + \lambda_n = 1, \quad 0 \leqslant \lambda_1, ..., \lambda_n \leqslant 1,$$
$$x_1, ..., x_n \in [a,b].$$

If g is twice differentiable, then we have a simple test to see if a function is convex, namely, g is convex if $g''(x) \geqslant 0$ for all $x \in [a, b]$. Geometrically one can show that if g is convex, then if we draw a line segment between any two points on the graph of the function, the entire segment lies above the graph of g, as we show formally bellow. A function g is *concave* if $-g$ is convex. Typical examples of convex functions are $g(x) = x^2$ and $g(x) = e^x$. Examples of concave functions are $g(x) = -x^2$ and $g(x) = \log x$. Convex and concave functions are always continuous.

Convex functions lie above tangents

Suppose $a < c < b$ and $g : [a, b] \longrightarrow \mathbb{R}$ be convex. Then there exist $A, B \in \mathbb{R}$ such that $g(c) = Ac + B$ and for all $x \in [a, b]$ we have $g(x) \geqslant Ax + B$.

PROOF. For $a \leqslant x < c < y \leqslant b$ we can write c as a convex combination of x and y, namely, $C = \lambda x + (1 - \lambda) y$ with $\lambda = \frac{y-c}{y-x} \in [0, 1]$. Therefore

$$g(c) \leqslant \lambda g(x)) + (1 - \lambda) g(y)$$

which implies that

$$\frac{g(c) - g(x)}{c - x} \leqslant \frac{g(y) - g(c)}{y - c}.$$

Thus we can take

$$\sup_{x<c} \frac{g(c) - g(x)}{c - x} \leqslant A \leqslant \inf_{y>c} \frac{g(y) - g(c)}{y - c},$$

so that we have for all $x < y$ in $[a, b]$ that

$$g(x) \geqslant A(x - c) + g(c) = Ax + (g(c) - Ac).$$

□

Proposition 15.7 (Jensen's inequality)

Suppose X is a random variable such that $\mathbb{P}(a \leqslant X \leqslant b) = 1$. If $g : \mathbb{R} \longrightarrow \mathbb{R}$ is convex on $[a, b]$, then

$$\mathbb{E}g(X) \geqslant g(\mathbb{E}X).$$

If g is concave, then

$$\mathbb{E}g(X) \leqslant g(\mathbb{E}X).$$

PROOF. If X is constant, then there is nothing to prove, so assume X is not constant. Then we have

$$a < \mathbb{E}X < b.$$

Denote $c := \mathbb{E}X$. Then

$$g(x) \geqslant AX + B \text{ and } g(\mathbb{E}X) = A\mathbb{E}X + B$$

for some $A, B \in \mathbb{R}$. Also note that

$$|g(X)| \leqslant |A||X| + |B| \leqslant |A \max\{|a|, |b|\}|X| + |B|,$$

so $\mathbb{E}|g(X)| < \infty$ and therefore $\mathbb{E}g(X)$ is well defined. Now we can use $AX + B \leqslant g(X)$ to see that

$$g(\mathbb{E}X) = A\mathbb{E}X + B = \mathbb{E}(AX + B) \leqslant \mathbb{E}g(X)$$

□

Example 15.8 (Arithmetic-geometric mean inequality)**.** Suppose a_1, ..., a_n are positive numbers, and X is a discrete random variable with the mass density

$$f_X(a_k) = \frac{1}{n} \text{ for } k = 1, ..., n.$$

Note that the function $g(x) = -\log x$ is a convex function on $(0, \infty)$. Jensen's inequality gives that

$$-\log\left(\frac{1}{n}\sum_{k=1}^{n} a_k\right) = -\log(\mathbb{E}X) \leqslant \mathbb{E}(-\log X) = -\frac{1}{n}\sum_{k=1}^{n}\log a_k.$$

Exponentiating this we get

$$\frac{1}{n}\sum_{k=1}^{n} a_k \geqslant \left(\prod_{k=1}^{n} a_k\right)^{1/n}.$$

Example 15.9. Suppose $p \geqslant 1$, then the function $g(x) = |x|^p$ is convex. Then

$$\mathbb{E}|X|^p \geqslant |\mathbb{E}X|^p$$

for any random variable X such that $\mathbb{E}X$ is defined. In particular,

$$\mathbb{E}X^2 \geqslant (\mathbb{E}X)^2,$$

and therefore $\mathbb{E}X^2 - (\mathbb{E}X)^2 \geqslant 0$.

15.7. Exercises

Exercise 15.1. Suppose we wire up a circuit containing a total of n connections. The probability of getting any one connection wrong is p. What can we say about the probability of wiring the circuit correctly? The circuit is wired correctly if all the connections are made correctly.

Exercise 15.2. Suppose $X \sim \operatorname{Exp}(\lambda)$. Using Markov's inequality estimate $\mathbb{P}(X \geqslant a)$ for $a > 0$ and compare it with the exact value of this probability.

Exercise 15.3. Suppose $X \sim \operatorname{Exp}(\lambda)$. Using Chebyshev's inequality estimate $\mathbb{P}(|X - \mathbb{E}X| \geqslant b)$ for $b > 0$.

Exercise 15.4. Suppose $X \sim \operatorname{Exp}(\lambda)$. Using Chernoff bounds estimate $\mathbb{P}(X \geqslant a)$ for $a > \mathbb{E}X$ and compare it with the exact value of this probability.

Exercise 15.5. Suppose $X > 0$ is a random variable such that $\operatorname{Var}(X) > 0$. Decide which of the two quantities is larger.

(A) $\mathbb{E}X^3$ or $(\mathbb{E}X)^3$?
(B) $\mathbb{E}X^{3/2}$ or $(\mathbb{E}X)^{3/2}$?
(C) $\mathbb{E}X^{2/3}$ or $(\mathbb{E}X)^{2/3}$?
(D) $\mathbb{E}\log(X+1)$ or $\log(\mathbb{E}X+1)$?
(E) $\mathbb{E}e^{X}$ or $e^{\mathbb{E}X}$?
(F) $\mathbb{E}e^{-X}$ or $e^{-\mathbb{E}X}$?

15.8. Selected solutions

Solution to Exercise 15.1: Let E_i denote the event that connection i is made correctly, so $\mathbb{P}(E_i^c) = p$. We do not anything beyond this (such as whether these events are dependent), so we will use Boole's inequality to estimate this probability as follows. The event we are interested in is $\cap_{i=1}^n E_i$.

$$\mathbb{P}\left(\bigcap_{i=1}^n E_i\right) = 1 - \mathbb{P}\left(\left(\bigcap_{i=1}^n E_i\right)^c\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^n E_i^c\right)$$
$$\overset{Boole}{\geqslant} 1 - \sum_{i=1}^n \mathbb{P}(E_i^c) = 1 - np.$$

Solution to Exercise 15.2: Markov's inequality gives

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}X}{a} = \frac{1}{a\lambda},$$

while the exact value is

$$\mathbb{P}(X \geqslant a) = \int_a^\infty \lambda e^{-\lambda x} dx = e^{-\lambda a} \leqslant \frac{1}{a\lambda}.$$

Solution to Exercise 15.3: We have $\mathbb{E}X = 1/\lambda$ and $\operatorname{Var} X = 1/\lambda^2$. By Chebyshev's inequality we have

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant b) \leqslant \frac{\operatorname{Var} X}{b^2} = \frac{1}{b^2\lambda^2}.$$

Solution to Exercise 15.4: recall first that

$$m_X(t) = \frac{\lambda}{\lambda - t} \text{ for } t < \lambda.$$

Using Chernoff bounds, we see

$$\mathbb{P}(X \geqslant a) \leqslant \min t > 0 \left(e^{-ta} m_X(t)\right) = \min_{t>0}\left(e^{-ta}\frac{\lambda}{\lambda - t}\right) \text{ for } t < \lambda.$$

To find the minimum of $e^{-ta}\frac{\lambda}{\lambda-t}$ as a function of t, we can find the critical point and see that it is $\lambda - 1/a > 0$ since we assume that $a > \mathbb{E}X = 1/\lambda$. Using this value for t we get

$$e^{-a\lambda} = \mathbb{P}(X \geqslant a) \leqslant a\lambda e^{1-a\lambda} = (a\lambda e) \cdot e^{-a\lambda} = (a\lambda e)\, \mathbb{P}(X \geqslant a).$$

Note that $a\lambda e \geqslant 1$.

Solution to Exercise 15.5(A): $\mathbb{E}X^3 > (\mathbb{E}X)^3$ since $(x^3)'' = x/3 > 0$ for $x > 0$.

Solution to Exercise 15.5(B): $\mathbb{E}X^{3/2} > (\mathbb{E}X)^{3/2}$ since $(x^{3/2})'' = \frac{3}{4\sqrt{x}} > 0$ for $x > 0$.

Solution to Exercise 15.5(C): $\mathbb{E}X^{2/3} < (\mathbb{E}X)^{2/3}$ since $(x^{2/3})'' = -\frac{2}{9x^{4/3}} < 0$ for $x > 0$.

Solution to Exercise 15.5(D): $\mathbb{E}\log(X+1) < \log(\mathbb{E}X+1)$ since $(\log(x))'' = -1/x^2 < 0$ for $x > 0$.

Solution to Exercise 15.5(E): $\mathbb{E}e^X > e^{\mathbb{E}X}$ since $(e^x)'' = e^x > 0$ for any x.

Solution to Exercise 15.5(F): $\mathbb{E}e^{-X} > e^{-\mathbb{E}X}$ since $(e^{-x})'' = e^{-x} > 0$ for any x.