1. **Convergence in Probability**

Let \((X_n)_{n \in \mathbb{N}}\) be a sequence of i.i.d. random variables distributed uniformly in \([-1, 1]\). Show that the following sequences \((Y_n)_{n \in \mathbb{N}}\) converge in probability to some limit.

a. \(Y_n = \prod_{i=1}^{n} X_i\).

b. \(Y_n = \max\{X_1, \ldots, X_n\}\).

c. \(Y_n = \frac{(X_1^2 + \cdots + X_n^2)}{n}\).

**Solution:**

a. By the independence of the random variables,

\[
\mathbb{E}(Y_n) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_n) = 0
\]

\[
\text{var}(Y_n) = \mathbb{E}(Y_n^2) = (\text{var}(X_1))^n = \left(\frac{1}{3}\right)^n.
\]

Since \(\text{var}(Y_n) \to 0\) as \(n \to \infty\), by Chebyshev’s Inequality, the sequence converges to its mean 0 in probability.

b. Consider \(\varepsilon \in [0, 1]\). We see that

\[
\mathbb{P}(|Y_n - 1| \geq \varepsilon) = \mathbb{P}(\max\{X_1, \ldots, X_n\} \leq 1 - \varepsilon)
\]

\[
= \mathbb{P}(X_1 \leq 1 - \varepsilon, \ldots, X_n \leq 1 - \varepsilon)
\]

\[
= \mathbb{P}(X_1 \leq 1 - \varepsilon)^n
\]

\[
= \left(1 - \frac{\varepsilon}{2}\right)^n,
\]

so \(\mathbb{P}(|Y_n - 1| \geq \varepsilon) \to 0\) as \(n \to \infty\), and we are done.

c. We can find the expectation, then bound the variance:

\[
\mathbb{E}(Y_n) = \frac{1}{n} \cdot n \mathbb{E}(X_1^2) = \frac{1}{3},
\]

\[
\text{var}(Y_n) = \frac{1}{n} \cdot \text{var}(X_1^2) \leq \frac{1}{n} \to 0 \quad \text{as } n \to \infty,
\]

since \(X_1^2 \leq 1\). Hence, we see that \(Y_n \to \frac{1}{3}\) in probability as \(n \to \infty\).
2. Really Random Binomial

You have a binomial random variable $X \sim \text{Binomial}(n, u)$. Unfortunately, you forget what the value of $u$ is, so you assume that $u$ is now a random variable $U \sim \text{Uniform}([0, 1])$, as you know that $0 \leq u \leq 1$. Given that you sample from this binomial distribution and observe $k$ successes, find the conditional distribution of $U$ given $X = k$.

**Hint:** Use MGFs to compute $\mathbb{P}(X = k)$ instead of integrating the distribution directly. The binomial theorem might also be useful here. Also, recall the identity $\sum_{i=0}^{n} s^i = \frac{1-s^{n+1}}{1-s}$.

**Solution:** By Bayes’ rule,

$$f_{U|X}(u | k) = \frac{p_{X|U}(k | u) \cdot f_{U}(u)}{\mathbb{P}(X = k)}.$$  

We know that $p_{X|U}(k | u) = \binom{n}{k} u^k (1-u)^{n-k}$ and $f_{U}(u) = 1$ by the distributions given. Finding the denominator by integrating over all values of $U$ is fairly tedious to do, as it requires iterative integration by parts, so we will instead resort to an approach based on MGFs.

$$M_X(s) = \mathbb{E}(e^{sX}) = \sum_{k=0}^{n} e^{sk} \mathbb{P}(X = k)$$

$$= \sum_{k=0}^{n} e^{sk} \int_{0}^{1} p_{X|U}(k | u) \cdot f_{U}(u) \, du$$

$$= \sum_{k=0}^{n} e^{sk} \int_{0}^{1} \binom{n}{k} u^k (1-u)^{n-k} \cdot 1 \, du$$

$$= \int_{0}^{1} \left( \sum_{k=0}^{n} \binom{n}{k} (u \cdot e^s)^k (1-u)^{n-k} \right) \, du$$

$$= \int_{0}^{1} (e^s u + (1-u))^n \, du.$$  

Note our use of the binomial theorem. Now we can substitute $v = (e^s u + (1-u))$, which gives $dv = -(1 - e^s) \, du$. Changing the limits appropriately,

$$= -\frac{1}{1-e^s} \int_{1}^{e^s} v^n \, dv$$

$$= 1 - e^{s(n+1)}$$

$$= \frac{1}{n+1} \sum_{k=0}^{n} e^{sk}.$$  

We observe that this MGF corresponds to a discrete random variable taking values in $0, \ldots, n$, each with probability $\frac{1}{n+1}$. In other words, $X$ is a uniform distribution over the specified range. Thus, our final posterior on $U$ is

$$f_{U|X}(u | k) = (n+1) \cdot \binom{n}{k} u^k (1-u)^{n-k}.$$  

**Remark:** Here is another way of understanding why $X$ is uniform. Recall that a binomial distribution can be written as the sum of i.i.d. Bernoulli random variables. Next, recognize
that if $V$ is uniform on $[0,1]$, the random variable $1_{V \leq u}$ is Bernoulli with parameter $u$. The difference is, the number $u$ is itself another uniform random variable, call it $U_{n+1}$. This means we can write $X = \sum_{i=1}^{n} 1_{U_i \leq U_{n+1}}$, and $X = k$ implies that $U_{n+1}$ occurs in the $(k+1)$-th position overall, which occurs with uniform probability $\frac{1}{n+1}$ since we are dealing with $n+1$ i.i.d. random variables.
3. Coupon Collector Bounds

Recall the coupon collector’s problem, in which there are \( n \) different types of coupons. Every box contains a single coupon, and we let the random variable \( X \) be the number of boxes bought until one of every type of coupon is obtained. The expected value of \( X \) is \( nH_n \), where \( H_n := \sum_{i=1}^{\infty} \frac{1}{i} \) is the harmonic number of order \( n \), which satisfies the inequality

\[
\ln n \leq H_n \leq \ln n + 1.
\]

a. Use Markov’s inequality in order to show that

\[
P(X > 2nH_n) \leq \frac{1}{2}.
\]

b. Use Chebyshev’s inequality in order to show that

\[
P(X > 2nH_n) \leq \frac{\pi^2}{6(\ln n)^2}.
\]

Note: You can use Euler’s solution to the Basel problem, the identity \( \sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6} \).

c. Define appropriate events and use the union bound in order to show that

\[
P(X > 2nH_n) \leq \frac{1}{n}.
\]

Note: \( a_n = (1 - \frac{1}{n})^n \) is a strictly increasing sequence with limit \( e^{-1} \).

Solution:

a. We are given \( \mathbb{E}(X) = nH_n \), so

\[
P(X > 2nH_n) \leq \frac{\mathbb{E}(X)}{2nH_n} = \frac{1}{2}.
\]

b. We can write \( X \) as an independent sum \( \sum_{i=1}^{\infty} X_i \), where \( X_i \sim \text{Geometric}(\frac{n-i+1}{n}) \), so

\[
\text{var}(X) = \sum_{i=1}^{n} \text{var}(X_i) < \sum_{i=1}^{n} \left( \frac{n}{n-i+1} \right) \left( \frac{n}{n-i+1} \right) = \sum_{i=1}^{n} \left( \frac{n}{i} \right)^2 < \frac{\pi^2 n^2}{6}.
\]

Using Chebyshev’s inequality, we have that

\[
P(X > 2nH_n) \leq \frac{\text{var}(X)}{(2nH_n)^2} < \frac{\pi^2}{6H_n^2} \leq \frac{\pi^2}{6(\ln n)^2}.
\]

c. Let \( A_i \) be the event that we fail to get box \( i \) after \( 2nH_n \) tries.

\[
P(A_i) \leq \left( \frac{n-i}{n} \right)^{2nH_n} = \left( \frac{n}{n-i+1} \right)^{nH_n} \leq e^{-2H_n} \leq e^{-2\ln n} = \frac{1}{n^2}.
\]

Now, by the union bound, we can conclude that

\[
P(X > 2nH_n) = P \left( \bigcup_{i=1}^{n} A_i \right) \leq \sum_{i=1}^{n} P(A_i) \leq \frac{1}{n}.
\]
4. Coupon Collector Convergence

In the coupon collector’s problem, there are \( n \) different types of coupons, and you are trying to collect them all. Each time you purchase an item, you receive one of the \( n \) coupons uniformly at random. Let \( T_n \) denote the number of purchases it takes to collect all \( n \) coupons. Prove that \( T_n/(n \ln n) \to 1 \) in probability as \( n \to \infty \).

**Solution:** We can write \( T_n \) as an independent sum \( \sum_{i=1}^{n} X_i \), where \( X_i \sim \text{Geometric}(\frac{n-i+1}{n}) \) is the number of purchases required to collect the \( i \)-th new coupon. Noting that the variance of a \( \text{Geometric}(p) \) random variable is \( \frac{1-p}{p^2} \leq \frac{1}{p^2} \), we have

\[
\mathbb{E}(T_n) = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \sum_{i=1}^{n} \frac{1}{i} = nH_n
\]

\[
\text{var}(T_n) = \sum_{i=1}^{n} \text{var}(X_i) \leq \sum_{i=1}^{n} \left( \frac{n}{n-i+1} \right)^2 = n \sum_{i=1}^{n} \left( \frac{n}{i} \right)^2 \leq n^2 \sum_{i=1}^{\infty} \frac{1}{i^2}.
\]

Here, it suffices to note that the summation converges. Notably,

\[
\text{var} \left( \frac{T_n - nH_n}{n \ln n} \right) \leq \frac{n^2}{n^2 (\ln n)^2} \sum_{i=1}^{\infty} \frac{1}{i^2} \to 0
\]

as \( n \to \infty \), and now Chebyshev’s inequality gives

\[
P \left( \left| \frac{T_n - nH_n}{n \ln n} \right| \geq \varepsilon \right) \leq \frac{1}{\varepsilon^2} \text{var} \left( \frac{T_n - nH_n}{n \ln n} \right) \to 0
\]

as \( n \to \infty \) for every \( \varepsilon > 0 \). Hence, \( (T_n - nH_n)/(n \ln n) \to 0 \) in probability as \( n \to \infty \). To conclude, we note that \( H_n \sim \ln n \) asymptotically, so \( T_n/(n \ln n) \to 1 \) in probability as \( n \to \infty \).

**Remark:** From previous analysis, we know that \( \mathbb{E}(T_n) \) is close to \( n \ln n \), so we have shown a result similar in spirit to a “weak law for the coupon collector problem”: as \( n \to \infty \), \( T_n \) is “close” to its expected value. However, since we are not dealing with i.i.d. random variables, we cannot use the version of the WLLN proved in lecture to deal with this problem.
5. The Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with common mean $\mu$ and MGF $M_X$. We assume that $M_X(s)$ is finite when $s \in (-d, d)$ for some $d > 0$. Let

$$
\bar{X}_n := \frac{X_1 + \ldots + X_n}{n}.
$$

a. Show that the transform (or MGF) associated with $\bar{X}_n$ satisfies

$$
M_{\bar{X}_n}(s) = M_X(s/n)^n.
$$

b. Suppose that the transform $M_X(s)$ has a first-order Taylor series expansion around $s = 0$ of the form

$$
M_X(s) = a + bs + o(s),
$$

where $o(s)$ is a function that satisfies $\lim_{s \to 0} o(s)/s = 0$. Find $a$ and $b$ in terms of $\mu$.

c. Show that for all $s \in (-d, d)$,

$$
\lim_{n \to \infty} M_{\bar{X}_n}(s) = e^{\mu s}.
$$

Hint: If $(a_n)_{n \in \mathbb{N}}$ is a sequence of real numbers converging to $a$, then $\lim_{n \to \infty} (1 + a_n/n)^n = e^a$.

d. Deduce that $\bar{X}_n \xrightarrow{d} \mu$. Note that the pointwise convergence of MGFs is equivalent to convergence in distribution.

**Solution:**

a. Let $S_n = X_1 + \cdots + X_n$. Then, as the sequence of random variables is i.i.d., we have that

$$
M_{S_n}(s) = M_X(s)^n.
$$

In addition,

$$
M_{\bar{X}_n}(s) = M_{S_n/n}(s) = \mathbb{E} \left( \exp \left( s \frac{S_n}{n} \right) \right) = M_{S_n}(s/n) = M_X(s/n)^n.
$$

b. $a = 1$ and $b = \mu$.

c. Taking the hint given,

$$
\lim_{n \to \infty} M_{\bar{X}_n}(s) = \lim_{n \to \infty} \left( 1 + \frac{\mu s + s o(1/n)}{n} \right)^n = e^{\mu s}.
$$

d. Let $Y = \mu$ be a constant random variable, and observe that $M_Y(s) = e^{\mu s}$. Since $M_{\bar{X}_n} \to M_Y$ as $n \to \infty$, as the MGF uniquely determines the distribution, we can deduce that $\bar{X}_n \xrightarrow{d} Y = \mu$ as $n \to \infty$. 
6. Decoding a Bit from a Noisy Signal

In many wireless communications systems, each receiver listens on a specific frequency. The bit $b$ sent is represented by a $+1$ or $-1$. Unfortunately, noise from other nearby frequencies can affect the receiver’s signal. A simplified model for this noise is as follows: There are $n$ other senders. The $i$th sender is also trying to send a bit $B_i$ represented by $+1$ or $-1$. The receiver obtains the signal $S$ given by

$$S = b + w \sum_{i=1}^{n} B_i,$$

where $w$ is a constant indicating the power of the bits of the other senders.

In order to decode $b$ from $S$, we use the following scheme: if $S$ is closer to $+1$ than $-1$, the receiver assumes that the bit sent was a $+1$; if $S$ is closer to $-1$ than $+1$, the receiver assumes that the bit sent was a $-1$; if $S$ is equidistant to $+1$ and $-1$, the receiver fails to recover $b$.

Assume that all the bits $B_i$ are independent and uniformly distributed over $\{+1, -1\}$.

a. Show that the probability that the receiver cannot determine $b$ correctly is at most $2 \exp(-\frac{1}{2nw^2})$. **Hint:** Transform each $B_i$ in order to use the following Chernoff bound, which you should show is true. If $X$ is Binomial($n, p$) as the sum of i.i.d. Bernoulli trials $X_1 + \cdots + X_n$, then

$$\mathbb{P}(X \geq nq) \leq \exp\left\{ -n \cdot (qt - \ln M_{X_1}(t)) \right\}$$

for all $t > 0$, and $\varepsilon > 0$ such that $q = p + \varepsilon < 1$. Then, you may use without proof the fact that for some value of $t$,

$$-n \cdot (qt - \ln M_{X_1}(t)) \leq -2n\varepsilon^2.$$

b. If we want to ensure that the probability to correctly determine $b$ is at least $1 - \delta = 0.999$, what condition do we need to impose on the power of the noise $w$?

c. What would be the condition on the power of the noise $w$ if we have used Chebyshev’s inequality in order to upper bound the error probability?

d. Discuss how the analysis of the error probability in part (a) compares with the analysis of the error probability using Chebyshev’s inequality.

Solution:

a. First, note that we get an error when $b = +1$ if $w \sum_{i=1}^{n} B_i \leq -1$, and we get an error when $b = -1$ if $w \sum_{i=1}^{n} B_i \geq 1$. Hence, the error probability is

$$\mathbb{P}\left(w \sum_{i=1}^{n} B_i \leq -1\right) = \mathbb{P}\left(w \sum_{i=1}^{n} B_i \geq 1\right) = \frac{1}{2} \cdot \mathbb{P}\left(\left|w \sum_{i=1}^{n} B_i\right| \geq 1\right),$$

where we also use the symmetry of the distribution around zero. In order to apply the Chernoff bound, we need to transform each $B_i$, which takes values in $\{-1, +1\}$, into a Bernoulli trial. Consider simply relabelling $-1$ to 0, i.e. the transformation $B'_i = \frac{B_i + 1}{2}$, under which $B_i \sim \text{Bernoulli}(\frac{1}{2})$. Now the error probability can be upper bounded by
\[
\mathbb{P}\left(w \sum_{i=1}^{n} B_i \geq 1\right) = \mathbb{P}\left(2 \sum_{i=1}^{n} B_i' - n \geq \frac{1}{w}\right) = \mathbb{P}\left(\sum_{i=1}^{n} B_i' - \frac{1}{2} n \geq \frac{1}{2w}\right).
\]

Let us show that the Chernoff bound given in the hint is true, letting \(q = p + \varepsilon\):

\[
\mathbb{P}(X \geq nq) \leq \frac{\mathbb{E}(e^{tX})}{e^{nqt}} = e^{-nqt} \prod_{i=1}^{n} \mathbb{E}(e^{tX_i}) = e^{-nqt} M_{X_1}(t)^n = \exp\{-n \cdot (qt - \ln M_{X_1}(t))\}.
\]

Finally, applying the result given with \(\varepsilon = \frac{1}{2nw}\) yields

\[
\mathbb{P}\left(w \sum_{i=1}^{n} B_i \geq 1\right) \leq \exp\left(-\frac{1}{2nw^2}\right).
\]

Note: It is fine if you upper bounded the probability

\[
\mathbb{P}\left(w \sum_{i=1}^{n} B_i \geq 1\right),
\]

as this will also give you an upper bound to the error probability as well, though loose by a factor of two.

b. To ensure that \(\mathbb{P}(w \sum_{i=1}^{n} B_i \geq 1) \leq \delta\), it suffices that \(\exp(-\frac{1}{2nw^2}) \leq \delta\), i.e. the power \(w\) needs to satisfy the condition

\[
 nw^2 \leq \frac{1}{2 \ln \frac{1}{\delta}} \approx 0.072.
\]

c. Using Chebyshev’s inequality, the error probability can be upper bounded by

\[
2 \cdot \mathbb{P}\left(w \sum_{i=1}^{n} B_i \geq 1\right) = \mathbb{P}\left(\left|w \sum_{i=1}^{n} B_i\right| \geq 1\right) \leq \text{var} \left(w \sum_{i=1}^{n} B_i\right) = nw^2.
\]

By this analysis, to ensure that \(\mathbb{P}(w \sum_{i=1}^{n} B_i \geq 1) \leq \delta\), we need the condition

\[
w^2 \leq 0.002.
\]

d. The conclusion we can draw is that in this setting, the analysis of the decoding procedure using the Chernoff bound for the binomial random variable is much tighter than the analysis using Chebyshev’s inequality. The decoding procedure can handle way more noise than what is indicated by Chebyshev’s inequality.