Ex (Coupon Collector's Problem) (Ev.2i).

What is the expected number of coupons one needs to collect before obtaining a complete set of all \( n \) types of coupons?

\[ X = X_0 + X_1 + \ldots + X_{n-1} \text{, where } X_i = \text{number of additional coupons (after } i \text{ types have been collected) in order to obtain a new type} \]

\[ E[X] = \sum_{i=0}^{n-1} E[X_i] \]

\( X_i \sim ? \)

When \( i \) types of coupons have already been collected, we are waiting for a new type. The coupons that we get now are of a new type with prob.

\[ p_i = \frac{n-i}{n} \Rightarrow \# \text{new types} \]

Hence

\[ X_i \sim \text{Geom}(p_i) \quad \Rightarrow \quad E[X_i] = \frac{1}{p_i} \]

\[ \Rightarrow \quad E[X] = \sum_{i=0}^{n-1} \frac{1}{p_i} = \frac{n}{n} + \frac{n}{n-1} + \ldots + \frac{n}{1} = n \left( 1 - \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \right) \]

Asymptotic analysis:

\[ 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \approx \ln n \] (harmonic series)

\[ \Rightarrow \quad E[X] \approx n \ln n \] (log - oversampling)

More precisely, harmonic number

\[ H_n = \sum_{k=1}^{n} \frac{1}{k} = \ln n + \gamma + \frac{1}{2} n^{-1} + O(n^{-2}) \]

\[ \Rightarrow \quad E[X] = n H_n + \gamma n + \frac{1}{2} + O(1) \]

Example: \( n = 50 \), \( E[X] = 22.5 \)

Remark: \( \text{Erdős - Renyi '61} \). \( P\{ X < n \ln n + \ln \} \to \exp(-e^t) \), \( n \to \infty \) \( (\forall t > 0) \)
Prop (Expectation of product) \[
\text{let } X, Y \text{ be independent r.v.'s. Then } \\
E(XY) = E(X) \cdot E(Y)
\]

Proof \[
E(XY) = \iint xy f(x,y) \, dx \, dy \quad \text{(by Prop. p. 87)} \\
= \left( \int_{-\infty}^{\infty} x f_X(x) \, dx \right) \left( \int_{-\infty}^{\infty} y f_Y(y) \, dy \right) \quad \text{(by independence)} \\
= E(X) \cdot E(Y). \quad \text{Q.E.D}
\]

Remark: independence is needed \[\begin{align*}
\text{Example: flip a coin. } & \quad H \Rightarrow X=1, Y=0 \\
& \Rightarrow X=0, Y=1 \\
\therefore E(XY) &= E(0) = 0 \\
\therefore E(X) &= E(Y) = \frac{1}{2}.
\end{align*}\]

Exercise: \[
\text{Estimate the average energy of a particle from the following data:} \\
\text{Average velocity of particle is } U_0, \text{ with st. dev. } \sigma_v; \\
\text{the mass of a particle is } m_0, \text{ with st. error of measurement } \sigma_m.
\]

\[
E = \frac{1}{2} m V^2. \quad \text{Ans: } \frac{1}{2} m_0 U_0^2 \quad \text{is somewhat incorrect.}
\]

\[
V = \text{velocity, } M = \text{mass, } E = \text{energy of a random particle}.
\]

\[
E = \frac{1}{2} M V^2. \quad \text{Under assumption that } M, V \text{ are independent,} \\
E[E] = \frac{1}{2} E[M] E[V^2].
\]

Given: \[
E[M] = m_0, \quad \text{but } E[V^2] \neq (E[V])^2 = U_0^2
\]

Instead: \[
\frac{E[V^2] - (E[V])^2}{U_0^2} = \frac{\text{Var}(V)}{U_0^2} = \sigma_v^2.
\]

\[
\Rightarrow E[V^2] = U_0^2 + \sigma_v^2.
\]

\[
\Rightarrow \text{Ans... } \quad E(E) = \frac{1}{2} m_0 (U_0^2 + \sigma_v^2).
\]
Alice and Bob want to cut out a rectangular piece of paper.

Alice guesses a random number $X$ and cuts out a square with sides $X$.

Bob guesses two random numbers $Y, Z$ (independent, and drawn from the same distribution)
and cuts out a rectangle with sides $Y, Z$.

Whose piece will have larger area, on average?

Alice's expected area is $E(x^2)$.

Bob's is $E(yz) = E(Y)E(Z)$ (by independence)

$= (E(x))^2$ (by identical distribution of $X$)

Since $E(x^2) - (E(x))^2 = Var(x) > 0$, Alice's piece will have larger area, on average.