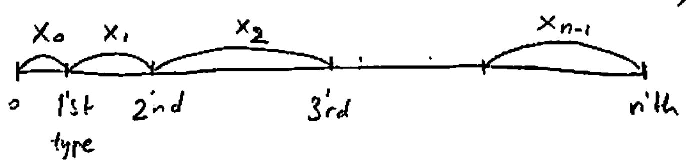
Ex (Coupon Collector's Problem II). (Ex. 2i)

What is the expected $\underbrace{\text{number of coupons}}_{X}$ one needs to collect before obtaining a complete set of all n types of coupons?

- $X = X_0 + X_1 + \dots + X_{n-1}$, where X_i = number of additional coupons (after i 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.

$$p_i = \frac{n-i}{n} \quad \begin{matrix} \leftarrow \text{\# new types} \\ \leftarrow \text{\# all types} \end{matrix}$$

Hence

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

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

- Asymptotic analysis (as $n \to \infty$). $1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \approx \ln n$ (harmonic series)

$$\Rightarrow \quad \boxed{E[X] \approx n \ln n} \qquad (\text{log-oversampling})$$

- More precisely, harmonic number $H_n = \sum_{k=1}^{n} \frac{1}{k} = \ln n + \gamma + \frac{1}{2} n^{-1} + o(n^{-2})$

$\gamma \approx 0.58$ (Euler-Mascheroni constant)

$$\Rightarrow \quad E[X] = n \ln n + \gamma n + \frac{1}{2} + o(1)$$

Example: $n = 50$, $E[X] \approx 225$

- **Remark**. [Erdös-Renyi '61]. $P\{X < n \ln n + tn\} \to \exp(-e^{-t})$, $n \to \infty$ $(\forall t > 0)$

Prop (Expectation of product) Let X, Y be independent r.v's. Then

$$E[X\cdot Y] = E[X]\cdot E[Y]$$

Proof
$$E(X\cdot Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, \underbrace{f(x,y)}_{\substack{\shortparallel \\ f_X(x)\, f_Y(y)\ \text{(by independence)}}}\, dx\, dy \qquad \text{(by Prop. p.87)}$$

$$= \left(\int_{-\infty}^{\infty} x f_X(x)\, dx\right)\left(\int_{-\infty}^{\infty} y f_Y(y)\, dy\right) = E[X]\cdot E[Y] \qquad \text{Q.E.D}$$

Remark. independence is needed. Example: flip a coin. $H \Rightarrow X=1, Y=0$; $T \Rightarrow X=0, Y=1$

$E[X\cdot Y] = E[0] = 0$

$E[X] = E[Y] = \frac{1}{2}$.

Ex Estimate the average energy of a particle from the following data:
Average velocity of particles is v_0, with st. deviation σ_v;
the mass of a particle is m_0, with st. error of measurement σ_m

$E = \frac{1}{2} m v^2$. Ans $= \frac{1}{2} m_0 v_0^2$ is somewhat incorrect.

V = velocity, M = mass, E = energy of a random particle.

$E = \frac{1}{2} M V^2$. Under assumption that M, V are independent,

$$E[E] = \frac{1}{2} E(M)\, E[V^2].$$

Given: $E[M] = m_0$, but $E[V^2] \underset{>}{\neq} (E[V])^2 = v_0^2$

Instead:
$$E[V^2] - (\underbrace{E[V]}_{\substack{\shortparallel \\ v_0^2}})^2 = Var(V) = \sigma_v^2.$$

$$\Rightarrow E[V^2] = v_0^2 + \sigma_v^2$$

$\Rightarrow$ Ans: $\boxed{E(E) = \frac{1}{2} m_0 (v_0^2 + \sigma_v^2)}$

Ex 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 taken from the same distr. as X) and cuts out a rectangle with sides Y, Z.

Whose piece will have larger area, on average?

Alices' expected area is $E(X^2)$.

Bob's is $E[YZ] = E(Y) \cdot E(Z)$ (by independence)

$= (E(X))^2$ (by identical distr. with X).

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