30/11 # LECTURE 18 : MORE ON RANDOM VARIABLES.

Eg 18.1 : (Coupon Collector problem) There are $n$ distinct coupons. Each day a coupon is selected indep. & UAR. All coupons are available for selection every day. (Sampling with replacement).

$T$ = # of days needed to get all $n$ distinct coupons. (r.v.)

Fix $m \geq 1$. $\mathbb{P}(T > m)$ = Prob. that in $m$ days, I haven't got $n$ distinct coupons.

$\mathbb{P}(T = m)$ = Prob. that I collect the $n$th distinct coupon on $m$th day.

$$\{T = m\} = \{T > m-1, T \leq m\} = \{T > m-1\} \cap \{T > m\}^c$$

$$\mathbb{P}(T = m) = \mathbb{P}(T > m-1) - \mathbb{P}(T > m)$$

$T > m$ - depends on the coupons received on the first '$m$' days.

$A_j$ = $j^{th}$ coupon isn't selected at all on the - '$m$' days.

$\{T > m\}$ = atleast one coupon isn't selected at all - - - '$m$' days

$$= \bigcup_{i=1}^{n} A_i$$

$$P(T>m) = P\left(\bigcup_{j=1}^{n} A_j\right) = S_1 - S_2 + S_3 \cdots + (-1)^{n+1} S_n$$

$$S_k = \sum_{i_1,\dots,i_k}^{+} P(A_{i_1} \cap \cdots \cap A_{i_k})$$

$$P(A_i) = P(i^{th} \text{ coupon isn't selected on day } 1, \dots, m)$$

(indep of selections) $$= \prod_{l=1}^{m} P(i^{th} \cdots \text{ on day } l)$$

$$= \prod_{l=1}^{m} \left(\frac{n-1}{n}\right) = \left(\frac{n-1}{n}\right)^m$$

$$P(A_i \cap A_j) = P(i^{th} \,\&\, j^{th} \text{ coupon aren't } \cdots \text{ on days } 1, \dots, m)$$

$$= \left(\frac{n-2}{n}\right)^m$$

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \left(\frac{n-k}{n}\right)^m$$

$$\Rightarrow S_k = \binom{n}{k}\left(\frac{n-k}{n}\right)^m$$

$$\mathbb{P}(T > m) = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k}\left(\frac{n-k}{n}\right)^m . \quad \blacksquare$$

Job of prob — Compute prob. $\longrightarrow$ Compute $\mathbb{E}[f(x)]$.

PROPN 18.2 (MARKOV'S INEQ.) Let $X$ be a r.v. & $X \geq 0$ (i.e., $p_X(x) = 0 \ \forall x < 0$.) then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t} \quad \forall t > 0.$$

[Check if $X(\omega) \geq 0 \ \forall \omega \in \Omega$ then $p_X(x) = 0 \ \forall x < 0$].

Proof: T.S.T.

$$t\,\mathbb{P}(X \geq t) \leq \mathbb{E}[X]$$

$$t\,\mathbb{E}[\mathbb{1}[X \geq t]] \leq \mathbb{E}[X]$$

$$\Updownarrow$$

linearity

$$\mathbb{E}[t\,\mathbb{1}[X \geq t]] \leq \mathbb{E}[X]$$

Observe that $t\,\mathbb{1}[X \geq t] \leq X \quad \forall\, t > 0.$

$$\left[\begin{array}{llll} X(\omega) \geq t & \Rightarrow \text{LHS} = t & ; & \text{RHS} = X(\omega) \ \& \ \text{LHS} \leq \text{RHS}. \\ X(\omega) < t & \Rightarrow \text{LHS} = 0 & ; & \text{RHS} = X(\omega) \geq 0 \ \& \ \text{LHS} \leq \text{RHS} \end{array}\right]$$

So by monotonicity of $\mathbb{E}$, we have that

$$\mathbb{E}[t\,\mathbb{1}[X \geq t]] \leq \mathbb{E}[X]$$

& so $\mathbb{P}(X \geq t) \leq \dfrac{\mathbb{E}[X]}{t} \quad \forall\, t > 0.$ ▨

**COR 18.3** Let $X$ be a r.v. & let $f : \mathbb{R} \to \mathbb{R}_+ = (0, \infty)$. & $f$ is non-decreasing.
Then $\forall\, t \in \mathbb{R}$

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E} f(X)}{f(t)} \quad \forall\, t \in \mathbb{R}.$$

**Proof:** If $X \geq t$ then $f(X) \geq f(t)$ as $f$ is non-decreasing.

So $\{X \geq t\} = \{\omega : X(\omega) \geq t\} \subseteq \{f(X) \geq f(t)\}$

By monotonicity, $\mathbb{P}(X \geq t) \leq \mathbb{P}(f(X) \geq f(t))$

(M. Ineq, as $f(X) \geq 0$) $\leq \dfrac{\mathbb{E}[f(X)]}{f(t)}$.

Eg 18.4 Let $f(x) = e^{\delta x}$, $\delta > 0$, $x \in \mathbb{R}$. $f(x) > 0$ & non-decreasing.

By above coro

$$\mathbb{P}(X \geq t) \leq e^{-\delta t}\,\mathbb{E}[e^{\delta X}] \quad \forall t \in \mathbb{R}.$$

(Chernoff's bound).

Ex 18.5 Let $X$ be a r.v. $\mathbb{P}(|X| \geq t) \leq \dfrac{\mathbb{E}[X^2]}{t^2}$ $\forall\, t > 0$.

Eg 18.6: $\Omega = S_n$, $\mathbb{P}(\sigma) = \dfrac{1}{n!}$ $\sigma \in S_n$.

Fix $k \in \{2, \ldots, n\}$

$\mathbb{P}(\exists$ a increasing subseq. of length $k) = ?$

Increasing subseq. of length $k$ is $i_1 < i_2 < \cdots < i_k$

$\ni \sigma(i_1) < \sigma(i_2) < \cdots < \sigma(i_k)$

For eg., $n=3$, $k=2$. $\sigma = (\sigma(1), \sigma(2), \sigma(3)) = (2,3,1)$.

only one incr. subseq. of length 2, $\sigma(1) < \sigma(2)$.

$\sigma = (1\,3\,2)$, 2 incr. subseq. of length 2 $\sigma(1)<\sigma(2)$

$\sigma = (1,2,3)$, 3 - - - of 2 $\sigma(1)<\sigma(3)$.

$X_k(\sigma) = \#$ of increasing subsequences of length $k$. (r.v.)

$\mathbb{P}(X_k \geq m) = ?$ i.e., Prob. $\exists$ $m$ increasing subsequences of length $k$.

$L(\sigma) =$ Length of longest increasing subsequence. (r.v.)

$= \max\{k : \exists$ an incr. subsequence of length $k\}$

$L: S_n \to [n]$ ; $X_k : S_n \longrightarrow \{0\} \cup \left[\binom{n}{k}\right]$

(check $X_k(\sigma) \leq \binom{n}{k} \ \forall \sigma \in S_n$)

$L(\sigma) = \max \{ k : X_k(\sigma) \geq 1 \}$

$\{ L(\sigma) \geq k \} = \{ X_k(\sigma) \geq 1 \}$.

$\{ L \geq k \} := \{ \sigma : L(\sigma) \geq k \}$ $\{ X_k \geq 1 \} = \{ \sigma : X_k(\sigma) \geq 1 \}$

$\mathbb{P}(L \geq k) = \mathbb{P}(X_k \geq 1)$

(Markov's ineq.) $\longrightarrow \leq \mathbb{E}[X_k]$.

$X_k(\sigma) = \# \{ (i_1, \ldots, i_k) : i_1 < \cdots < i_k \ \& \ \sigma(i_1) < \cdots < \sigma(i_k) \}$

$= \sum_{i_1 < \cdots < i_k} \mathbb{1}[\sigma(i_1) < \cdots < \sigma(i_k)]$

$$X_{i_1,\dots,i_k}(\sigma) = \mathbb{1}[\sigma(i_1) < \dots < \sigma(i_k)]$$

Use linearity $\mathbb{E}[X_k] = \sum_{i_1<\dots<i_k} \mathbb{E}[X_{i_1,\dots,i_k}]$

(check) $= \sum_{i_1<\dots<i_k} \mathbb{P}(\{\sigma : \sigma(i_1) < \dots < \sigma(i_k)\})$

$$= \sum_{i_1<\dots<i_k} \frac{|\{\sigma : \sigma(i_1) < \dots < \sigma(i_k)\}|}{n!}$$

(check) $= \sum_{i_1<\dots<i_k} \frac{1}{k!} = \binom{n}{k}\frac{1}{k!}$

$\mathbb{P}(L \geq k) \leq \binom{n}{k}\frac{1}{k!}$ Find seq. $k_n \ni$ RHS $\to 0$ as $n \to \infty$.

Dan Romik — "The beautiful mathematics of LIS".

Exe Use union bound to compute $\mathbb{P}(X_k \geq 1)$.

Ex we claim work to compute $\mathbb{E}[X_k] = \ldots$

Defn 18.7 Define variance of X ($\mathrm{VAR}[X]$) by

$$\mathrm{VAR}[X] := \mathbb{E}\left[ \left( X - \mathbb{E}[X] \right)^2 \right]$$

(STD. DEVN.) $$SD[X] := \sqrt{\mathrm{VAR}[X]}$$

OBS: $$\mathrm{VAR}[X] = \mathbb{E}[X^2] - \left( \mathbb{E}[X] \right)^2$$

$$\mathrm{VAR}[X] = \mathbb{E}\left[ \left( X - \mathbb{E}[X] \right)^2 \right] = \mathbb{E}\left[ \left( X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2 \right) \right]$$

(lin. of exp.) $$= \mathbb{E}[X^2] - 2\mathbb{E}\left[ X\,\mathbb{E}[X] \right] + \mathbb{E}\left[ \left( \mathbb{E}[X] \right)^2 \right]$$

linearity

$$= \mathbb{E}[X^2] - 2\,\mathbb{E}[X]\,\mathbb{E}[X] + \left( \mathbb{E}[X] \right)^2$$

$$= \mathbb{E}[X^2] - \left( \mathbb{E}[X] \right)^2.$$ ∎

Ex. If X is const r.v. (i.e. $P_X(a) = 1$ for some $a \in \mathbb{R}$) then $\mathbb{E}[X] = a$.