Regarding Problem 4

X conditioned on $U=u$ r.v. has mean $\sqrt{u}$ variance u call it X_u

U r.v. density $f_U(u)$

$$\mathrm{Var}(Y) = E(Y^2) - (E(Y))^2$$

$$\Rightarrow E(Y^2) = \mathrm{Var}(Y) + (E(Y))^2$$

$$E(X) = \int_{-\infty}^{\infty} \overbrace{E(X \mid U=u)}^{E(X_u)} f_U(u)\,du$$

$$= \int_{-\infty}^{\infty} \sqrt{u}\, f_U(u)\,du \quad \text{Suppose} = 5$$

$$E(X^2) = \int_{-\infty}^{\infty} \underbrace{E(X^2 \mid U=u)}_{\underbrace{E(X_u^2)}_{\mathrm{Var}(X_u) + (\mathrm{mean}(X_u))^2}} f_U(u)\,du$$

$$\cancel{\mathrm{Var}(X) = \int_{-\infty}^{\infty} \mathrm{Var}(X \mid U=u) f_U(u)\,du}$$

$$\mathrm{Var}(X) = E\Big(\underbrace{(X - E(X))^2}_{W}\Big)$$

$$E(W) = \int_{-\infty}^{\infty} \underbrace{E(W \mid U=u)}_{\cancel{\mathrm{Var}(X_u)}} f_U(u)\,du$$

$$E(W \mid U=u) = E\big((X-5)^2 \mid U=u\big) = E\big((X_u - 5)^2\big)$$

Streaming Algs

see elts going by, very little space to store anything.

$a_1, a_2, \ldots, a_T \ldots$ each $a_i \in U$

Distinct Elements

a_1	a_2	a_3	$\ldots$								
32	5	17	32	14	5	17	5	17	5	32	17

$Y_1(t)$: 0.43 0.43 0.19 0.19 ... 0.19

$Y_2(t)$ ...

$Y_k(t)$ ...

$\bar{Y}(t)$

$h: U \to (0,1)$

$h_1(32) = 0.43$

$h_1(5) = 0.61$

$h_1(17) = 0.19$

$h_1(14) = 0.85$

$h: U \to [m]$

$h(x) = i$

think of it as mapping to $\frac{i}{m}$

Define $Y(t) = \min_{1 \le j \le t} h(a_j)$ storing one hash value only

$$\boxed{E[Y(t)] = \frac{1}{n_t + 1}}$$

n_t # of distinct elts in $a_1, a_2, \ldots, a_t$

$Y(t) = \frac{1}{n_t+1} \Rightarrow n_t + 1 = \frac{1}{Y(t)}$

estimator for n_t $\rightarrow$ $\boxed{\hat{n}_t = \frac{1}{Y(t)} - 1}$

Chebyshev:

$\Pr(|X - E(X)| \ge c\sigma) \le \frac{1}{c^2}$

$\Pr\left(\left|Y(t) - \frac{1}{n_t+1}\right| \ge c \frac{1}{n_t+1}\right) \le \frac{1}{c^2}$

0 $\frac{1}{n+1}$

$\mathrm{Var}(Y(t)) \approx \frac{1}{(n_t+1)^2}$

Consider k indep hash fns $h_1, \ldots, h_k$

$1 \le i \le k$ $Y_i(t) = \min_{1 \le j \le t} h_i(a_j)$

$\bar{Y}(t) = \frac{Y_1(t) + \ldots + Y_k(t)}{k}$ $E(\bar{Y}(t)) = \frac{1}{n_t+1}$

$\mathrm{Var}(\bar{Y}(t)) = \frac{1}{k} \mathrm{Var}(Y(t))$ $\sigma(\bar{Y}(t)) = \frac{\sigma(Y(t))}{\sqrt{k}}$

$$Pr\left(\left|\bar{Y}(t) - \frac{1}{n_t+1}\right| \geq 2\,\frac{1}{\sqrt{k}\,(n_t+1)}\right) \leq \frac{1}{4}$$

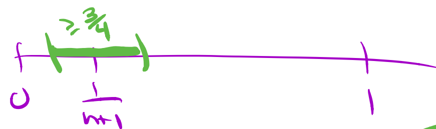
$k = 16$

$$Pr\left(\left|\bar{Y}(t) - \frac{1}{n_t+1}\right| > \frac{2}{4(n_t+1)}\right) \leq \frac{1}{4}$$

$$\frac{1}{2(n_t+1)}$$

w.p. $\geq \frac{3}{4}$

$$\frac{1}{n+1} - \frac{1}{2(n+1)} \leq \bar{Y}(t) \leq \frac{1}{n+1} + \frac{1}{2(n+1)}$$

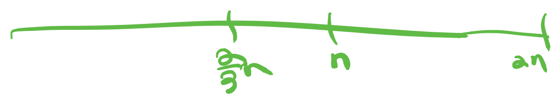


$$\frac{1}{2(n+1)} \leq \bar{Y}(t) \leq \frac{3}{2(n+1)}$$

$$\boxed{\hat{n} = \frac{1}{\bar{Y}(t)} - 1}$$

$$\frac{2}{3}(n+1) \leq \frac{1}{\bar{Y}(t)} \leq 2(n+1)$$

$$\frac{2}{3}(n+1) - 1 \leq \underbrace{\frac{1}{\bar{Y}(t)} - 1}_{\hat{n}} \leq 2(n+1) - 1$$



$$E(g(x)) \neq g(E(x))$$

$$E\left(\frac{1}{x}\right) \neq \frac{1}{E(x)}$$

$Y = \min(X_1, \ldots, X_n)$

$X_i \sim U[0,1]$

X_i's are iid.

$E(Y)$

$F_Y(y) = Pr(Y \leq y)$

$= Pr(\min(X_1, \ldots, X_n) \leq y)$

$= 1 - Pr(\min(X_1, \ldots, X_n) > y)$

$= 1 - Pr(X_1 > y, X_2 > y, \ldots, X_n > y)$

$= 1 - (1 - F_X(y))^n$

for $U(0,1)$ $= 1 - (1-y)^n$

$f_Y(y) = n(1-y)^{n-1}$

$$E(Y) = \int_0^1 y \, n(1-y)^{n-1} \, dy$$