Coupon Collector

Jan 28
Eva Tardos
et 28

Geometric variable

$Geom(p)$ $\quad prob(H) = p$ $\qquad Geom(p) = \#$ times till H

X is $Geom(p)$

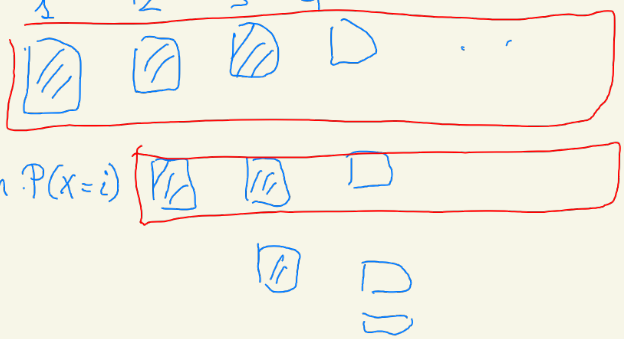
(1) $\quad E(X) = \sum_{n=0}^{\infty} n \Pr(X=n)$

$\Pr(X=n) = (1-p)^{n-1} p$

Claim $\quad X \geq 0$ integer $\qquad E(X) = \sum_{n=1}^{\infty} \Pr(X \geq n)$

Proof: $\quad x \quad 1 \quad 2 \quad 3 \quad 4 \ldots$

$\Pr(x=1$



$\Pr(X \geq 1)$

$n \cdot P(X=i)$

$\Pr(X \geq 2)$

version (1) add columns

version (2) add rows

$$E(X) = \sum_{n=1}^{\infty} n \Pr(X=n) = \sum_{n=1}^{\infty} \sum_{k=1}^{n} \Pr(X=n) = \sum_{k=1}^{\infty} \Pr(X \geq k)$$

reverse sum

X geom(p)

$$E(X) = \sum_{n=1}^{\infty} Pr(X \geq n) = \sum_{k=1}^{\infty} (1-p)^{k-1} = \frac{1}{1-(1-p)} = \frac{1}{p}$$

Different method

$$X = \begin{cases} 1 & \text{prob } p \quad \text{if first flip} = H \\ 1 + Y & \text{" } (1-p) \end{cases}$$

Y: # flip needed after the first

principle of defered decision

$$E(X) = p \cdot 1 + (1-p)\, E(1+Y) \underset{\text{linearity}}{=} p \cdot 1 + (1-p) \cdot 1 + (1-p)\, E(Y)$$

$E(X) = E(Y)$ using this $E(X) = 1 + (1-p)\, E(X)$

$$\Rightarrow E(X) = \frac{1}{p}$$

Variance of Geom(p)

$$Var(X) = E((X - E(X))^2) = E(X^2) - 2E(X\,E(X)) + E(X)^2$$

$$2E(X\,E(X)) = 2\,E(X)^2$$

linearity of exp
$E(X)$ above a number

$$= E(X^2) - E(X)^2, \quad E(X)^2 = \frac{1}{p^2}$$

using second method

$$E(X^2) = p\,1 + (1-p)\, E((1+Y)^2)$$

$$= p \cdot 1 + (1-p)\left[E(1) + 2E(Y) + E(Y^2)\right] \quad (X \,\&\, Y \text{ same})$$

$$= p + (1-p)\left(1 + \frac{2}{p} + E(X^2)\right)$$

$$E(X^2)\,p = 1 + \frac{1-p}{p} \longrightarrow E(X^2) = \frac{2-p}{p^2}$$

$$V(X^2) = E(X^2) - E(X)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

Coupon collector

n coupons possible

each pick each coupon equally likely

τ = time to collect all n coupons

$E(\tau)$? τ_k time till get k different coupons

$\tau_1 = 1$. $X_{k+1} = \tau_{k+1} - \tau_k$ time waiting for $k+1$ st coupon

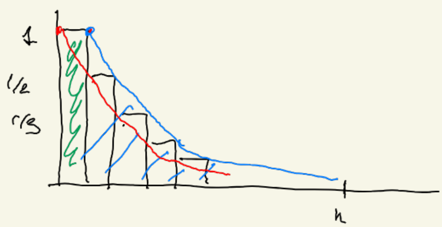
Claim X_{k+1} is geom $\left(\frac{n-k}{n}\right)$

Using this
$\tau = X_1 + X_2 + \cdots + X_n$ using linearity

$$E(\tau) = \sum_{k=0}^{n-1} E(X_{k+1}) = \sum_{k=0}^{n-1} \frac{n}{n-k} = n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{1}\right)$$

Fact $\sum_{k=1}^{n} \frac{1}{k} \sim \ln n$

Claim



consider $\frac{1}{x} = f(x)$

$$\sum_{k=1}^{n} \frac{1}{k} \le 1 + \int_1^n \frac{1}{x}\,dx = 1 + [\ln x]_1^n = 1 + \ln n$$

consider $g(x) = \frac{1}{x+1}$

$$\sum_{k=1}^{n} \frac{1}{k} \ge \int_0^{n-1} \frac{1}{x+1}\,dx = \ln n$$

so $E(\tau) \sim n \ln n$ almost exactly

Variance of τ

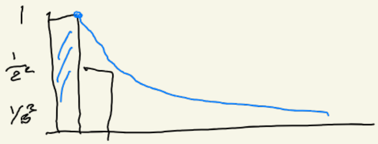
$\tau = X_1 + \cdots + X_n$

Claim X_i are independent

$$\Longrightarrow V(\tau) = V(X_1) + \cdots + V(X_n) = \sum_{k=0}^{n-1} \frac{1 - \frac{n-k}{n}}{\left(\frac{n-k}{n}\right)^2}$$

recall X_{k+1} geom $\frac{n-k}{n}$ $Var(X_{k+1}) = \frac{1 - \frac{n-k}{k}}{\left(\frac{n-k}{n}\right)^2}$

$$\le \sum_{k=0}^{n-1} \frac{1}{\left(\frac{n-k}{n}\right)^2} = n^2 \sum_{k=0}^{n-1} \frac{1}{(n-k)^2} = n^2\left(\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\dots\frac{1}{n^2}\right) \le 2n^2$$



consider $f(x) = \frac{1}{x^2}$

$$\sum_{k=1}^{n} \frac{1}{k} \le 1 + \int_1^n \frac{1}{x^2}\,dx = 1 + \left[-\frac{1}{x}\right]_1^n = 1 - \frac{1}{n} + 1$$
$$= 2 - \frac{1}{n}$$

---

m ($\approx$ time that make $\Pr(c \le m) \le \frac{1}{2}$

Chebyshev: X $\Pr(|X - E(X)| \ge \alpha) \le \frac{Var(X)}{\alpha^2}$

what α makes prob $\le 1/2$

$$\frac{2n^2}{\alpha^2} \le \frac{1}{2} \qquad \text{e.g. } \alpha = 2n$$

$\Pr(c > E(c) + 2n) \le 1/2$

$m = n\log n + 2n$ makes prob $\ge 1/2$