CS 70 ## 4.9 LECTURE 22

Coupon Collection

There are n coupons. Our goal is to collect all. If L is the # of cereal boxes I buy to obtain all n distinct coupons, what's $E(L)$?

The first is always unique, so start at second. $(n-1)/n$ 1, $1-\frac{n-1}{n}$ 0

$L_1 = 1$, $L_2 \sim \text{Geom}\left(\frac{n-1}{n}\right)$, $L_j \sim \text{Geom}\left(\frac{n-j+1}{n}\right)$

$L = L_1 + L_2 + \ldots + L_n \longrightarrow E(L) = \sum_{j=1}^{n} E(L_j) = \sum_{j=1}^{n} \frac{1}{p_j} = \sum_{j=1}^{n} \frac{n}{n-j+1}$ $\quad k = j-1$

$E(L) = \sum_{k=0}^{n-1} \frac{n}{n-k} = n \sum_{k=0}^{n-1} \frac{1}{n-k} = n\underbrace{\left(\frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{2} + 1\right)}_{\text{harmonic sum } H_n = \sum_{\ell=1}^{n} \frac{1}{\ell} \approx \ln n + \gamma + \frac{1}{2n}} = n \ln n + n\gamma + \frac{1}{2}$

gamma

$\gamma = 0.5772$, and this is called Euler's constant.

If $n=5$, $E(L) = 5 \ln 5 + 5\gamma + \frac{1}{2} = 11.42 \longrightarrow 12$ boxes

Variance

$\sigma_X^2 = \text{Var}(X) = E\left[(X - E(X))^2\right] \geq 0 \longrightarrow$ standard deviation $\sigma_X = \sqrt{\text{Var}(X)}$

The units of σ are the same as X, and σ^2 are X^2.

$\sigma_X^2 = E\left[X^2 - 2X E(X) + E^2(X)\right] = \underbrace{E(X^2)}_{2^{nd}\text{ moment of } X} + E(-2X E(X)) + \underbrace{E(E^2(X))}_{E^2(X)} = E(X^2) - 2E^2(X) + E^2(X)$

$= E(X^2) - E^2(X)$

If X is a zero-mean RV (ie. $E(X)=0$), $\sigma_X^2 = E(X^2)$. When is $\sigma_X^2 = 0$? iff $X = E(X)$

EX What is σ_X^2 for a Bernoulli RV? $\underbrace{E(K^2)}_{\sum_{K=0}^{1} k^2 P_K(k)} - \underbrace{E^2(K)}_{p^2} = p - p^2 = p(1-p)$

What kind of a coin would give you maximum (un)certainty?

For certainty, we'd want $p=1$ so $\sigma^2 = 0$. For uncertainty, $p = 0.5$.

Poisson RV

Say you have a book with $n >> 1$ words, and the probability of each word being a typo is $0 < p << 1$. Let L be the # of typos in the book.

Generally speaking, a large # of Bernoulli trials with low prob. of success.

$P_L(\ell) = \binom{n}{\ell} p^{\ell} (1-p)^{n-\ell} \quad \forall \ell \in \{0, \ldots, n\}$ $\qquad E(L) = \lambda \rightarrow$ "average arrival rate" $= np$

CLAIM As $n \to \infty$, $P_L(\ell) = \lambda^{\ell}/\ell! \cdot e^{-\lambda}$

PROOF $P_L(\ell) = \frac{n!}{\ell!(n-\ell)!} \left(\frac{\lambda}{n}\right)^{\ell} \left(1-\frac{\lambda}{n}\right)^{n} \Big/ \left(1-\frac{\lambda}{n}\right)^{\ell} = \frac{n(n-1)\cdots(n-\ell+1)}{\ell!\, n^{\ell}} \lambda^{\ell} \left(1-\frac{\lambda}{n}\right)^{n-\ell}$ $\quad p = \lambda/n$

$= \lambda^{\ell}/\ell! \cdot \underbrace{\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-\ell+1}{n}}_{1 \cdot 1 \cdots 1} \underbrace{\left(1-\frac{\lambda}{n}\right)^{n-\ell}}_{e^{-\lambda}}$ $\qquad$ from calculus, $\lim_{x\to\infty}(1+x)^{1/x} = \lim_{y\to\infty}(1+1/y)^{y} = e$

$= \lambda^{\ell}/\ell! \cdot e^{-\lambda}$

$\blacksquare$

Mean of Poisson RV

$$E(L) = \sum_{\ell=0}^{\infty} \ell \frac{\lambda^\ell}{\ell!} e^{-\lambda} = \lambda$$

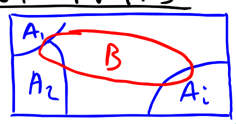
we want $\lambda = np$

$$= \sum \frac{\lambda^\ell}{(\ell-1)!} e^{-\lambda} = \lambda \left[\sum \frac{\lambda^\ell}{(\ell-1)!}\right] e^{-\lambda} = \lambda \cdot e^{\lambda} \cdot e^{-\lambda} = \lambda$$

Variance of Poisson RV

Also λ.

Joint PMFs



$$P(B) = \sum_{i=1}^{\infty} P(B \cap A_i)$$

Marginal PMF $P_X(x) = P(X=x)$

$$P_{X,Y}(x,y) = P_r(X=x \cap Y=y)$$

$$P_{X|Y}(x|y) = \frac{P_r(X=x \cap Y=y)}{P_Y(y)} = \frac{P_{X,Y}(x,y)}{P_Y(y)}$$

Plotting this in 2D is hard. Anything higher is hopeless.