Elements of Probability Summer 2019

Birthday Paradox and random variables — 23.03, 2019

prof. dr hab. Jacek Cichoń *Scribe: Krzysztof Agieńczuk*

1 Overview

In the last lecture we talked about independence and conditional probability.

In this lecture we covered:

- birthday paradox,
- Bayes formula,
- random variables,
- geometric distribution.

2 Birthday Paradox

Let's say we have n urns and k balls, which we want to put into those urns. By collision we will mean an event in which two balls fall into one urn. As urns are identical, probability of choosing an urn is $\frac{1}{n}$.
$\Omega = \{1, ..., n\}^{\{1,...,k\}}$
$|\Omega_{n,k}| = n^k$
We are looking for probability $B_{n,k}$ of throwing two balls into one urn.
$B_{n,k} = \{(x_1, ..., x_n) \in \Omega_{n,k} : (\forall 1 \leq i < j \leq k)(x_i \neq x_j)\}$
$P(B_{n,k}) = \frac{|B_{n,k}|}{n^k}$
There are n places for the first ball. n-1 for second, n-2 for third and n-(k-1) for k-th.
$P(B_{n,k}) = \frac{n(n-1)...(n-(k-1))}{n*n*...*n} = (1 - \frac{1}{n})(1 - \frac{2}{n}) * ... * (1 - \frac{k-1}{n})$

$$P(B_{n,k}) = \Pi_{a=1}^{k}(1 - \frac{a}{n}) \tag{1}$$

Now, this function looks similar to e^x. Also, we know for a fact, that $e^x \geq 1 + x$, so we can say
$\Pi_{a=1}^{k}(1 - \frac{a}{n}) \leq \Pi_{a=1}^{k} e^{-\frac{a}{n}} = e^{\sum_{a=1}^{k-1}(-\frac{a}{n})} = e^{-\frac{1}{n}\frac{k(k-1)}{2}}$
When $-\frac{1}{n}\frac{k(k-1)}{2} = -1 \equiv k^1 = 2n \equiv k = \sqrt{2n}$
Then

$$P[B_{n,k}] \simeq \frac{1}{e} \simeq \frac{1}{3} \tag{2}$$

Example Birthday paradox as is (ie. n=365 days in a year).
$\sqrt{2 * 365} \simeq \sqrt{700} \simeq 26$

Note $\sqrt{2n}$ is a point after which it is very likely that there are two balls in one urn. There is another point, $n * ln(n)$ at which we are sure there is a ball in every urn.

This problem occurs in hashing functions. We need very large numbers, because collisions are very possible in quickly.

3 Bayes formula

We know conditional probability denoted as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{3}$$

Let's multiply it by P(A). We then get

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)} \tag{4}$$

which we know as Bayes formula. Exemplary usage in Big Data are bayesian networks.

Example Let's say we have test predicting if people are suffering from certain disease. The test manufactures say that it is 99% accurate, which means $P(T|S) = 0.99$ (probability of test being True, when Sick). This implies $P(T|H) = 0.01$. Let's assume probability of being sick is $P(S) = 0.001$ (which is not unreasonable).
$P(S|T) = P(T|S)\frac{P(S)}{P(T)} = P(T|S)\frac{P(S)}{P(T \cap S)+P(T \cap H)} = P(T|S)\frac{P(S)}{P(T|S)P(S)+P(T|H)P(H)}$
After substituting data:
$P(S|T) = 0.99\frac{10^{-3}}{(1-10^{-2})10^{-3}+10^{-2}(1-10^{-3})} * \frac{10^3}{10^3} = 0.99\frac{1}{(1-10^{-2})+10^{-2}(10^3-1)} \approx \frac{1}{11} \approx 0.1$
Which shows that probability of actually being sick when the test was positive is small. This implies a lot of false positives.

4 Random variables

$(\Omega, \mathcal{S}, P)$ is a probability space. $X : \Omega \to \mathbb{R}$.

Definition 4.1. *Random variable: X is a random variable ($\mathcal{S}$-measurable) if*

$$(\forall a \in \mathbb{R})(X^{-1}((a, \infty)) \in \mathcal{S}) \tag{5}$$

This means
$X^{-1}((a, \infty)) = \{\omega \in \Omega : X(\omega) \in (a, \infty)\} = \{\omega \in \Omega : X(\omega) > a\}$

Fact

1. X is a random variable

2. $(\forall a < b)(X^{-1}((a,b)) \in \mathcal{S})$
3. $(\forall a < b)(X^{-1}([a,b]) \in \mathcal{S})$
4. $(\forall B \in Bor(\mathbb{R}))(X^{-1}(B) \in \mathcal{S})$[1]

4.1 Discrete Probability Spaces

Definition 4.2. *For discrete probability space* $\Omega = \{\omega_1, ..., \omega_n\}$
$\mathcal{S} = P(\Omega)$
$E(X) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) = \sum_{i=1}^{n} X(\omega)P(\{\omega\})$

Example Suppose $P(\{\omega_i\}) = \frac{1}{n}$
$E(X) = \sum_{\omega \in \Omega} X(\omega)\frac{1}{n} = \frac{1}{n}\sum_{\omega \in \Omega} X(\omega) = \frac{1}{n}\sum_{k=1}^{n} X(\omega_k)$

Let's assume $X : \Omega \to \mathbb{R} \quad range(X) = \{a_1, ..., a_n\}$
$E(X) = \sum_{\omega} X(\omega)P(\{\omega\}) = \sum_{i=1}^{k}\sum_{\omega} X(\omega)X(\omega)P(\{\omega\}) = \sum_{i=1}^{k} a_i \sum_{\omega \in \Omega; \omega = a_i} P(\{\omega\}) =$
$E(X) = \sum_{i=1}^{k} a_i P(\{\omega \in \Omega : X(\omega) = a_i\})$ which gives us

$$E(X) = \sum_{a \in Rng(X)} aP(X = a) \tag{6}$$

5 Geometric distribution

We define p as "probability of success" in range $p \in [0, 1]$. And q is defined as $q = 1 - p$. Let's imagine we throw a coin and p is probability of throwing a head. For L number of steps probabilities go as follows
$P[L = 1] = p$
$P[L = 2] = qp$
$P[L = 3] = q^2 p$
and so on. For L=k
$P[L = k] = pq^{k-1}$

Definition 5.1. $L \sim Geo(p) \equiv range(L) \subseteq \{1, 2, ...\} \wedge (\forall k > 1)P(L = k) = pq^{k-1}$

Assume $L \sim Geo(p)$

$$E(L) = \sum_{k=1}^{\infty} kP(L = k) = \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty}(kq^{k-1}) \tag{7}$$

$$\sum_{k>0}^{\infty} q^k = \frac{1}{1-q} for |q| < 1 \tag{8}$$

[1] X^{-1} means preimage of a function

$$\sum_{k\geq} x^k = \frac{1}{1-x} \tag{9}$$

$$\sum_{k\geq 1} kx^{k-1} = \frac{1}{(1-x)^2} for |x| < 1 \tag{10}$$

If we substitute p and q we get

$$E(L) = p\frac{1}{(1-p)^2} = \frac{p}{p^2} = \frac{1}{p} \tag{11}$$

Fact $(L \sim Geo(p)) \implies E(L) = \frac{1}{p}$