Probability Theory & Statistics

Innopolis University, BS-I,II

Spring Semester 2016-17

Lecturer: Nikolay Shilov

Part I

MEAN, AVERAGE, EXPECTATION, THE FIRST MOMENT

Definition

- Let $X:\Omega\rightarrow R$ be a random variable and P_X be its probability distribution.
- *Mean*, *average*, *expectation*, the *first moment* of X is defined as

$$M(X) = E(X) = \Sigma_{x\in R}x*P_X(x).$$

- Is this definition correct?

Examples

- Rolling the dice: since $P_X(x)=1/6$ for all x in {1, 2, 3, 4, 5, 6}, then $M(X)= 1*1/6 + 2*1/6 + 3*1/6 + 4*1/6 + 5*1/6 + 6*1/6 = 7/2$.
- Lottery prize (ref. week 5): since

Prize	1000	100	1	0
Probability	0.0001	0.001	0.01	0.9889

$M(X)= 1000*0.0001 + 100*0.001 + 1*0.01 + 0*0.9889 = 0.1 + 0.1 + 0.01 = 0.21$.

Two exercises

- What is the average sum of pips
  - on an domino tile?
  - on pair of dices?

Linearity of the expectation

- Prove: if $a \in R$ is a constant then
  - $M(a) = a$;
  - $M(a*X) = a*M(X)$;
- Discuss: if X and Y have the same outcomes then $M(X + Y) = M(X) + M(Y)$;
- Prove: if $a,b \in R$ are constants and X and Y have the same outcomes then

$$M(a*X + b*Y) = a*M(X) + b*M(Y).$$

Expectation of a function

- Let $X:\Omega\to R$ be a random variable and $g:R\to R$ be function (that is defined on the range of X at least).
- Prove: if $Y=g(X)$ then $M(Y)= \Sigma_{x\in R} g(x)*P_X(x)$.

Independent Random Variables

- Random variables $X,Y:\Omega\to R$ are said to be *independent* if $(X=x)$ and $(Y=y)$ are independent events for all $x,y\in R$.
- Prove: random variables $X,Y:\Omega\to R$ are independent iff $P(X=x \cap Y=y) = P(X=x)*P(Y=y)$ for all $x,y\in R$.
- Question: what the probability space is used in the definition and property?

Expectation of Product of Two Independent Random Variables

- Prove: if random variables $X,Y:\Omega\rightarrow R$ are independent then $M(X*Y) = M(X)*M(Y)$.
- Question: is equality $M(X*Y) = M(X)*M(Y)$ valid for all random variables $X,Y:\Omega\rightarrow R$? (Either prove or provide a counterexample.)

Part II

SELECTED DISCRETE DISTRIBUTIONS

Discrete Uniform Distribution

- Discrete uniform distribution corresponds to a random variable X that get exactly n values $\{x_1, \ldots, x_n\}$ with a flat probability:

$$P_X(x_k)=1/n \text{ for all } k\in[1..n].$$

- Expectation $M(X) = (x_1 + \ldots + x_n)/n$ is the (arithmetic) mean of X values.
- Example: tossing a dice.

Bernoulli Trial and Distribution

- X = Bernoulli(p), where p$\in$[0, 1], gets just 2 conventional values 0 (fail) and 1 (success).
- Bernoulli distribution:
  - P(Bernoulli(p)=1) = p and
  - P(Bernoulli(p)=0) = 1-p.
- M(Bernoulli(p)) = 0*(1-p) + 1*p = p.

Bernoulli Trials and Binomial Distribution

- Bernoulli trials (or binomial experiment) consist of some fixed number n of independent Bernoulli trials, each with a probability of success p, and counts the number of successes.
- A corresponding random variable is denoted by B(n,p) or by binomial(n,p), and is said to have a binomial distribution:

$$P_{B(n,p)}(k) = P(B(n,p)=k) = C_n^k p^k (1-p)^{(n-k)}.$$

Mean for Binomial Distribution

- Let X=B(n,p).
- By expectation definition:

$$M(X)= \Sigma_{k\in[0..n]}k*C_n^k p^k(1-p)^{(n-k)}.$$

- By definition of B(n,p): $X = X_1 + \ldots + X_n$ where all $X_1,\ldots X_n$ =Bernoulli(p).
- By expectation linearity: M(X) = n*p.
- Intuitively: number of success in n trials should be n*p.

Random Variable with Geometric Distribution

- The probability distribution of the number k in { 1, 2, 3, ...} of Bernoulli(p) trials needed to get the first success:

$$P(X=k) = (1-p)^{(k-1)} * p.$$

- Attention: for the first time the range of *possible* (i.e. not impossible) values is *infinite*; so the set of outcomes Ω must be infinite too.

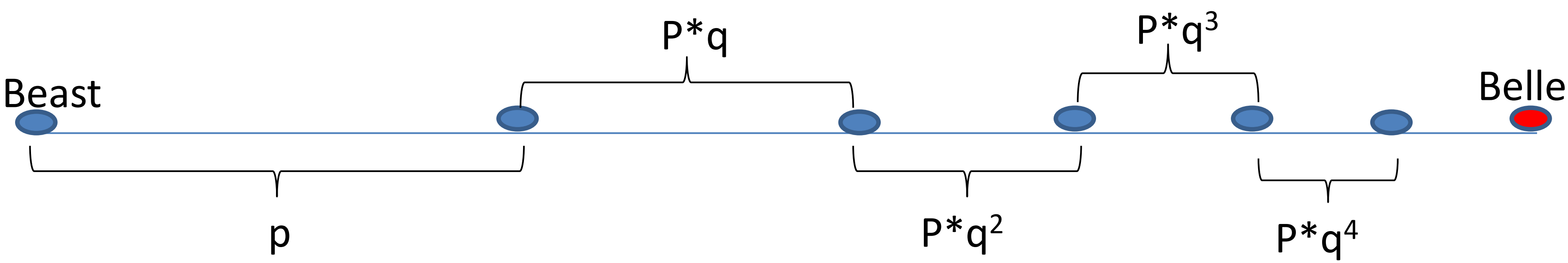
Example: Beauty and the Beast

The Beast is a shy guy: he stays just 1 meter behind Belle … but can’t dare to call her.

http://kinoprofi.org/7041-krasavica-i-chudovische-chudesnoe-rozhdestvo-1997.html

Example: Beauty and the Beast (cont.)

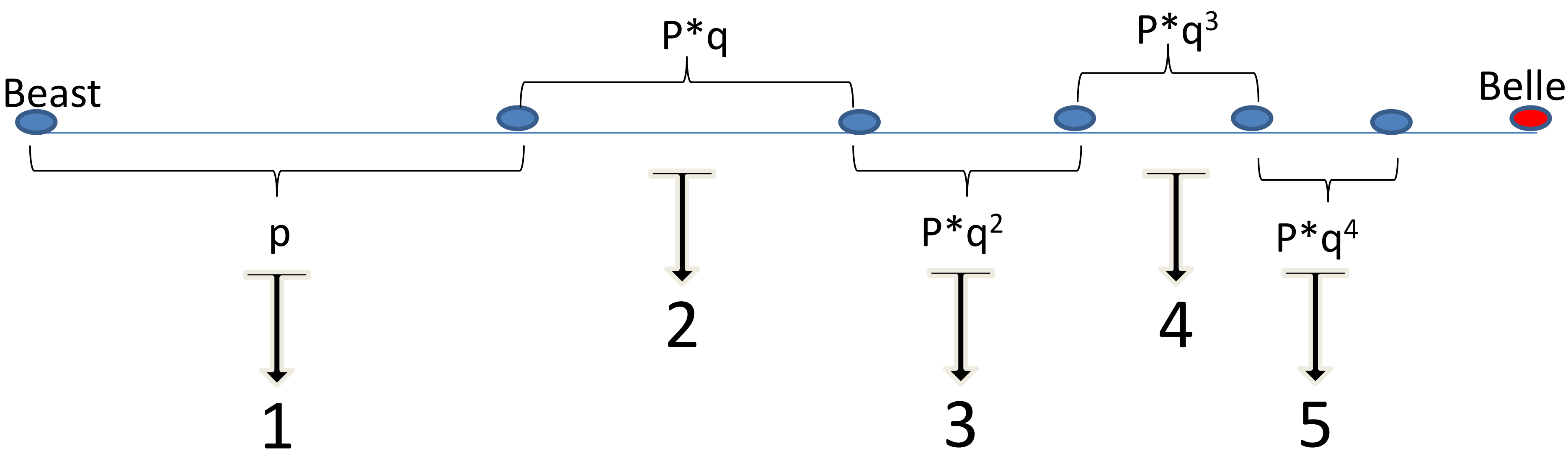
- So he makes the first step to her of p cm length and (maybe) some more steps so that each next step is q=(1-p/100cm) smaller then the previous one.



- After that the Beast calls Belle. What is probability that he calls after (exactly) k steps?

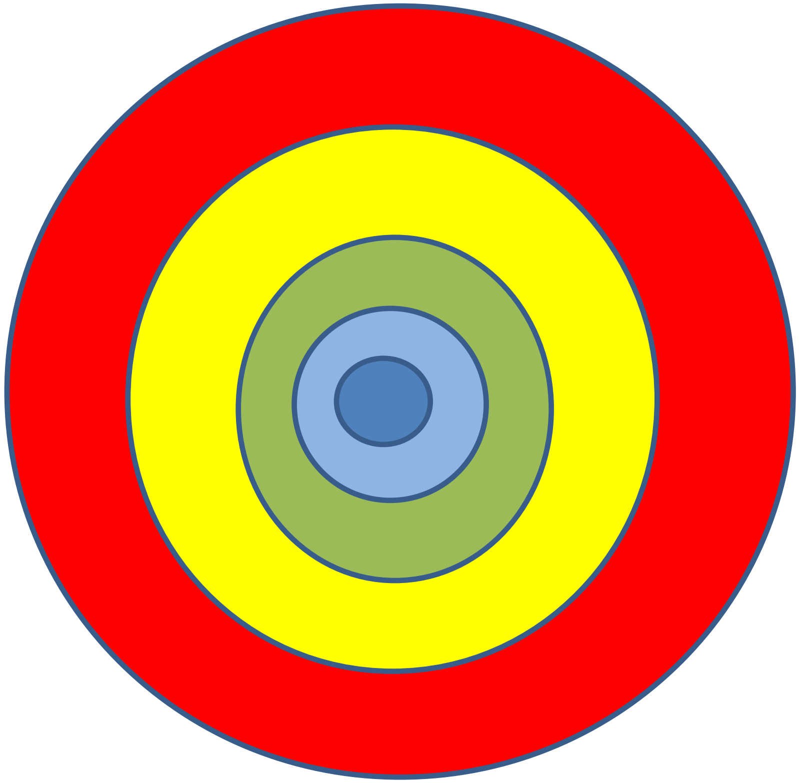
Example: Beauty and the Beast (cont.)

- Let X:[0, 1)$\rightarrow$R be the following staircase function: X(t)=k on interval $[(1\text{-}q^{(k-1)}), (1\text{-}q^{k}))$.
- Ω=[0,1) and X= Geom(p):



Build another example of X=Geom(p)

Find radiuses of rings such that areas of concentric rings X=Geom(p):



Mean for Geometric Distribution

- If X=Geom(p) then

$$M(X) = \Sigma_{k>0}\, k*p*(1-p)^{(k-1)} = 1/p$$

(refer the next part for a proof).

Part III

CONDITIONAL AND TOTAL EXPECTATION

Recall form Lecture for week 6

- Discrete random variable is any (total) real function on finite domain $X:\Omega\rightarrow R$.
- (Discrete) probability distribution is a table/function that assigns probability (

$$P_X(x) = P(X=x) = \frac{|X^{-}(x)|}{|\Omega|}$$

of the corresponding outcomes to each value x in the range of the variable X.

Recall form Lecture for week 6 (cont.)

- There are several ways how to affiliate a probability space with random variable $X:\Omega\to R$.
- One particular way:
  - $\Omega' = \{ S\subseteq\Omega : S=X^{-}(x), x\in R \}$;
  - $\mathcal{F}= 2^{\Omega'}$;
  - $P:\mathcal{F}\to[0,1]$ is the additive continuation on $\mathcal{F}$ of a function defined as $P(X^{-}(x))=P_X(x)$ for every $x\in R$.

Other way to Probability Space

- Adopt the set of outputs Ω as the sample space,
- use the standard event space $\mathcal{F}= 2^{\Omega}$;
- expand (in additive way) onto $\mathcal{F}$ the standard probability function $P(\omega)=1/|\Omega|$ for $\omega\in\Omega$.

Conditional (Discrete) Random Variable (simple case)

- Assume $A \subseteq \Omega$ is an “event” in 2^{Ω}. Then *conditional random variable* $X|A$ is the restriction of the function X on the domain A.
- It defines *conditional distribution*

  $P_{X|A}(x) = P((X|A)=x) =$

  $= |(X=x) \cap A|/|A| = |X^{-}(x) \cap A|/|A| =$

  $= P((X=x)|A).$
- Corollary: $P((X=x) \cap A) = P_{X|A}(x)*P(A)$

Conditional Expectation

- Conditional expectation – expectation of conditional random variable:

$$M(X|A) = \Sigma_{x\in R} x * P_{X|A}(x) .$$

- If $H_1, \ldots H_n$ is a partition then

$$P(A) = \Sigma_{k\in[1..n]} P(A|H_k) * P(H_k)$$

and hence

$$P_X(x) = \Sigma_{k\in[1..n]} P_{X|Hk}(x) * P(H_k).$$

Total Expectation Formula

$$M(X) = \Sigma_{x\in R}x*P_X(x) =$$

$$= \Sigma_{x\in R}x*(\Sigma_{k\in[1..n]}P_{X|Hk}(x)*P(H_k)) =$$

$$= \Sigma_{x\in R}\Sigma_{k\in[1..n]}\ x*P_{X|Hk}(x)*P(H_k) =$$

$$= \Sigma_{k\in[1..n]}\Sigma_{x\in R}\ x*P_{X|Hk}(x)*P(H_k) =$$

$$= \Sigma_{k\in[1..n]}(\Sigma_{x\in R}\ x*P_{X|Hk}(x))*P(H_k) =$$

$$= \Sigma_{k\in[1..n]}M(X|H_k)*P(H_k)$$

Part IV

EXAMPLES OF CONDITIONAL AND TOTAL EXPECTATION

Back to Lottery (week 5)

Prize	1000	100	1	0
Probability	0.0001	0.001	0.01	0.9889

- If all tickets are sold out then M(X)= 1000*0.0001 + 100*0.001 + 1*0.01 + 0*0.9889 = 0.1 + 0.1 + 0.01 = 0.21.
- The minimal circulation consists of 10000 tickets including 1 with prize 1000, 10 - with prize 100 each, and 100 – with prize 1 each.

Back to Lottery (cont.)

- Event A: just 1000 tickets are sold out from the minimal circulation.
- $M(X|A) = \Sigma_{x\in R} x^* P_{X|A}(x) =$

  $= 1000 * 1/1000 * P(1000\text{-ticket is sold}) +$

  $+ 100 * 1/1000 * P(1 \text{ of } 100\text{-tickets is sold}) + \ldots$

  $+ 100 * 10/1000 * P(10 \text{ of } 100\text{-tickets are sold}) +$

  $+ 1 * 1/1000 * P(1 \text{ of } 1\text{-tickets is sold}) + \ldots$

  $+ 1 * 100/1000 * P(10 \text{ of } 100\text{-tickets are sold}) =$

  $= 1000/9001 + \ldots > 0.111 + \ldots > 0.1 + \ldots = 0.21.$

Memoryless Property

- *Memorylessness* refers to the cases when the distribution of a waiting until a certain event does not depend on time.
- Only two kinds of distributions are memoryless: exponential and geometric distributions.

Memorylessness of Geometrical Distribution

- Let
  - $X=$ Geom(p) and
  - $(X>m) = \{\omega \in \Omega : X(\omega)>m\}$.
- $P(X=(m+n) \mid X>m) = P(X= m+n)/P(X>m) = P(n)$

Mean for Geometric Distribution

- If X=Geom(p) then $M(X) = \Sigma_{k>0}\, k*p*(1-p)^{(k-1)}$.
- Using total expectation formula with
  - $H_1 = (X=1)$, $P(H_1)=p$,
  - $H_2 = (X>1)$, $P(H_2)=q = (1-p)$:

  $M(X) = M(X|H_1)*P(H_1) + M(X|H_2)*P(H_2)$.

Mean for Geometric Distribution (cont.)

- Due to memorylessness:

  $P(X=(1+n) \mid X>1) = P(X=n)$ for all $n>0$;
- hence $M(X|H_2) = M(1 + X) = 1 + M(X)$;
- It implies that $M(X) = 1*p + (1 + M(X))*q$;
- hence $M(X)= 1/p$.