NMAI059 Probability and statistics 1
Class 4

Robert Šámal

Overview

Discrete r.v. – expectation and variance

Parameters of discrete distributions

Random vectors

What we have learned

- What is a discrete r.v.
- How to describe it using a PMF and/or CDF.
- Examples of distributions: Bernoulli, binomial, hypergeometric, Poisson, geometric.
- Expectation: two possible definitions
- $\mathbb{E}(X) = \sum_{x \in Im(X)} x \cdot P(X = x)$
- $\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$
- $\mathbb{E}(g(X)) = \sum_{x \in Im(X)} g(x) P(X = x)$ (LOTUS)
- "How much we expect to get on average, when we repeat independent experiments with result given by X" … we will discuss later as the law of large numbers.

$X : \Omega \to \mathbb{R}$

$P(X = x)$

$P(X \le x)$

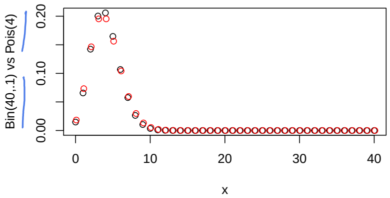
X — time for a quicksort algo
run it N-times
total running time

$N \cdot \mathbb{E}(X) = \sum_{x \in Im(X)} x \cdot \underbrace{P(X=x) \cdot N}_{\text{\# of attempts that end up with time } X = x}$

we expect

expected total running time

Comparing binomial and Poisson distribution: PMF



Generated by the following code in R

```
x = 0:40
bin = dbinom(x,40,0.1)
pois = dpois(x,4)
plot(x,bin, ylab="Bin(40,.1) vs Pois(4)")
points(x+.1,pois,col="red")
```

Properties of $\mathbb{E}$

$P(X<0)=0 \Rightarrow \forall x<0 \quad P(X=x)=0$

$\textcircled{3}\ E(aX+b) \overset{LOTUS}{=} \sum_{x\in \mathrm{Im} X} (ax+b)P(X=x) = a\underbrace{\sum x\, P(X=x)}_{EX} + b\sum 1.\, P(X=x)$

Theorem

Suppose X, Y are discrete r.v. and $a, b \in \mathbb{R}$.

1. *If $P(X \geq 0) = 1$ and $\mathbb{E}(X) = 0$, then $P(X = 0) = 1$.*
2. *If $\mathbb{E}(X) \geq 0$ then $P(X \geq 0) > 0$.*
3. $\mathbb{E}(a \cdot X + b) = a \cdot \mathbb{E}(X) + b$.
4. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

Proof $\textcircled{1}$ $\underset{0}{EX} = \sum_{x\in \mathrm{Im}(X)} x \cdot P(X=x) = \underset{0}{0 \cdot P(X=0)} + \underbrace{\sum_{\substack{x\in \mathrm{Im}(X)\\ x>0}} x\, \overset{0}{P(X=x)}}_{0} + \underbrace{\sum_{\substack{x\in \mathrm{Im}(X)\\ x<0}} x\, \overset{0}{P(X=x)}}_{=0}$

$\Rightarrow P(X=x)=0 \quad \forall x>0$

$\textcircled{2}$ $0 \leq EX = \sum_{x\in \mathrm{Im}(X)} x\, P(X=x) = 0 \cdot \overset{=0}{P(X=0)} + \sum_{x>0} x\, \overset{0}{P(X=x)} + \underbrace{\sum_{x<0} x\, P(X=x)}_{<0}$

If, for the sake of contradiction, $P(X\geq 0)=0$ then $\ne 0$ for some x

$\Rightarrow$ contradiction !!

(4) $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ linearity of expectation

in case Ω is discrete!

$$\mathbb{E}(X+Y) = \sum_{\omega\in\Omega} \big(X(\omega)+Y(\omega)\big)\cdot P(\{\omega\})$$

$$= \underbrace{\sum X(\omega)P(\{\omega\})}_{\mathbb{E}(X)} + \underbrace{\sum Y(\omega)P(\{\omega\})}_{"\mathbb{E}(Y)}$$

Another formula for expectation

Theorem

Let X be a discrete r.v. such that $Im(X) \subseteq \mathbb{N}_0 = \underline{\{0, 1, 2, \dots\}}$. Then we have

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} P(X > n).$$

(application: geom. distr.)

Proof

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k P(X = k)$$

$$= \sum_{k=0}^{\infty} \sum_{n=0}^{k-1} P(X = k) \qquad n < k$$

$$P(\{\omega : X(\omega) = k\})$$

$$= \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} P(X = k)$$

$$\{\omega : X(\omega) > n\}$$

$$= \sum_{n=0}^{\infty} P\Big(\underbrace{\bigcup_{k=n+1}^{\infty} \{\omega : X(\omega) = k\}}\Big) = \sum_{n=0}^{\infty} P(X > n)$$

$\blacksquare$

Variance

$0 \quad 1 \quad 2 \quad \mathbb{E}X \quad 10$

Definition

Variance of a r.v. X *is the number* $\mathbb{E}((X-\mathbb{E}X)^2)$*. It is denoted by* $var(X)$*.*

$\mathbb{E}X^2 \neq (\mathbb{E}X)^2$

Theorem

$\boxed{\mu = \mathbb{E}X}$

$$var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

$\Rightarrow \mathbb{E}(X^2) \geq (\mathbb{E}X)^2$

$\mathbb{E}(X^2) = (\mathbb{E}X)^2 \iff var\,X = 0 \iff X = \text{const. a.s.}$

Proof $var(X) = \mathbb{E}((X-\mu)^2) = \mathbb{E}(X^2 - 2\mu X + \mu^2)$

$= \mathbb{E}(X^2) - 2\mu \underbrace{\mathbb{E}(X)}_{\mu} + \mu^2$

$= \mathbb{E}(X^2) - \mu^2$

standard deviation

$\sigma_X = \sqrt{var(X)}$

Conditional expectation

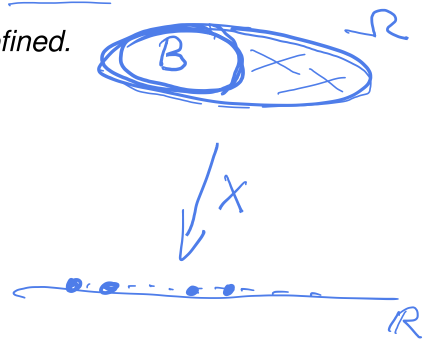
Definition

Let X be a discrete r.v. and $P(B) > 0$. Conditional expectation of X given B is

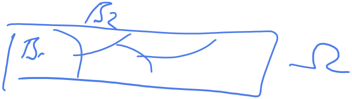
$$\mathbb{E}(X \mid B) = \sum_{x \in Im(X)} x \cdot P(X = x \mid B),$$

whenever the sum is defined.

expectation of X restricted to B



Law Of Total Expectation



Theorem

Suppose $B_1, B_2, \ldots$ is a partition of Ω and ~~$A \in \mathcal{F}$~~ X is a discrete r.v.. Then

$$\mathbb{E}(X) = \sum_i P(B_i)\mathbb{E}(X \mid B_i),$$

whenever the sum is defined. (Terms with $P(B_i) = 0$ are counted as 0.)

Proof

$$\mathbb{E}(X) \overset{?}{=} \sum_i P(B_i) \cdot \mathbb{E}(X \mid B_i)$$

$$= \sum_i P(B_i) \sum_{x \in Im(X)} x \cdot P(X = x \mid B_i)$$

$$= \sum_{x \in Im(X)} x \underbrace{\sum_i P(B_i) P(X = x \mid B_i)}_{= P(X = x)} = \sum x \cdot P(X = x) \quad \blacksquare$$

Law Of Total Expectation

$X \sim Geom(p)$ $\qquad p =$ prob. of success

$X =$ time till success
"roll a 6"

$p = \frac{1}{6}$

$\Omega = \{1,2,\dots,6\}^{\mathbb{N}}$

$S =$ "first coordinate $= 6$"

$X =$ "index of first 6"

$B_1 = S =$ success in the first roll

$B_2 = F =$ fail — " —

$$E X = P(S)\,\underline{E(X|S)} + P(F)\cdot\underline{E(X|F)}$$

$$P(S) = p, \quad E(X|S) = 1, \quad P(F)\cdot E(X|F) = (1-p)(1+E(X))$$

$\omega = (2,4,6,\dots) \in F$

$X(\omega) = 3$

$$\underline{E(X)} = p\cdot 1 + (1-p)(\underline{E(X)} + 1)$$

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

$p = \frac{1}{6} \dots E(X) = 6$

Overview

Discrete r.v. – expectation and variance

Parameters of discrete distributions

Random vectors

Distribution parameters – Bernoulli

Pro $X \sim Bern(p)$ je

- $\mathbb{E}(X) = p$ ✓
- $var(X) = p - p^2 = p(1-p)$ ✓

$X = \begin{cases} 1 & \text{w. prob. } p \\ 0 & 1-p \end{cases}$

$\underline{\mathbb{E}X} = 1 \cdot P(X=1) + 0 \cdot P(X=0) = \underline{P(X=1)} = p$

$var(X) = \mathbb{E}(X-p)^2 = (1-p)^2 \cdot P(X=1) + (0-p)^2 \cdot P(X=0)$

$\boxed{\text{Yes}}$ $= (1-p)^2 p + p^2(1-p) = p(1-p) \cancel{(p+1-p)}$

$var(X) = \underline{\mathbb{E}}(X^2) - (\mathbb{E}X)^2 = p - p^2$ $\boxed{\text{No}}$

$X = X^2$!

Distribution parameters – binomial

$[A] = I_A$:= r.v. that is $= 1$ if A, $= 0$ if not A

For ~~Pro~~ $X \sim Bin(n, p)$ ~~je~~ we have

- $\mathbb{E}(X) = np$
- $var(X) = np(1-p)$

- First way: $\boxed{X = \sum_{i=1}^{n} X_i,}$ where $X_i = [i\text{-th attempt was a success}]$
- $\mathbb{E}(X_i) = P(X_i = 1) = p$

$EX = \sum_{i=1}^{n} EX_i = np$ (Linearity of exp.)

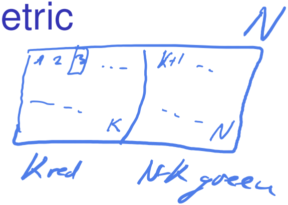
- Second way:
  $\mathbb{E}(X) = \sum_{k=0}^{n} k \cdot P(X = k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$

$$= \sum_{k=1}^{n} pn \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}$$

$$= pn \left(p + (1-p)\right)^{n-1} = np$$

Distribution parameters – hypergeometric

Pro $X \sim Hyper(N, K, n)$ = # of red out of n balls drawn w/o rept.



- $\mathbb{E}(X) = n\frac{K}{N}$
- $var(X) = n\frac{K}{N}(1 - \frac{K}{N})\frac{N-n}{N-1}$

- First way: $X = \sum_{i=1}^{n} X_i$, where $X_i = [i\text{-th ball is red}]$
- $\mathbb{E}(X_i) = P(X_i = 1) = \frac{K}{N}$ --- obvi for $i=1$

? $i \geq 1$ .... don't look at the first $i-1$ balls, too the i-th one ... $\frac{K \text{ good cases}}{N \text{ total}}$

$\mathbb{E}X = n \cdot \mathbb{E}(X_i) = n \cdot \frac{K}{N}$

- Second way: $X = \sum_{j=1}^{K} Y_j$, where $Y_j =$
- $\mathbb{E}(Y_j) = P(Y_j = 1) = \frac{n}{N}$

$Y_j = [$we draw ball with id. j in n attempts$]$

$$P(Y_j = 1) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

$$\mathbb{E}X = \sum_{1}^{K} \mathbb{E}Y_j = K \cdot \frac{n}{N}$$

put balls in a row

our balls

in a random order ($N!$ ways)

Distribution parameters – geometric

For $X \sim Geom(p)$ we have

- $\mathbb{E}(X) = 1/p$ ✓
- $var(X) = \frac{1-p}{p^2}$

Law of t. Exp. ✓

$$\mathbb{E}X = \sum_{n=0}^{\infty} P(X > n)$$

$$P(X > n) = (1-p)^n \implies \mathbb{E}X = \sum_{n=0}^{\infty} (1-p)^n = \frac{1}{1-(1-p)} = \frac{1}{p}$$

$$\sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \cdot p = \ldots\ldots$$

Distribution parameters – Poisson

Pro $X \sim Pois(\lambda)$ je

- $\mathbb{E}(X) = \lambda$
- $var(X) = \lambda$

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \underbrace{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda}}_{=1} = \lambda$$

$g(x) = x(x-1)$

LOTUS

$\mathbb{E}(X^2 - X)$

$$\mathbb{E}(X(X-1)) = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} = \lambda^2$$

$\mathbb{E}X^2 - \mathbb{E}X$

$$var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \underbrace{(\mathbb{E}X^2 - \mathbb{E}X)}_{\lambda^2} + \underbrace{\mathbb{E}X}_{\lambda} - \underbrace{(\mathbb{E}X)^2}_{\lambda^2} = \lambda$$

Overview

Discrete r.v. – expectation and variance

Parameters of discrete distributions

Random vectors

Basic description of random vectors

- X, Y – random variables on the same probability space $(\Omega, \mathcal{F}, P)$.
- We wish to treat (X, Y) as one object – a random vector.
- How to do that?
- Example: we roll twice a 4-sided dice, $X =$ first outcome, $Y =$ second one.

$X \backslash Y$	1	2	3	4
1	$\frac{1}{16}$	?	?	0 ?
2				
3				
4				

Joint distribution

Definition

For a discrete r.v. X, Y *on a probability space* $(\Omega, \mathcal{F}, P)$ *we define their joint PMF* $p_{X,Y} : \mathbb{R}^2 \to [0,1]$ *by a formula*

$$p_{X,Y}(x,y) = P(\{\omega \in \Omega : X(\omega) = x \& Y(\omega) = y).$$

- We can define it also for more than two r.v.'s $p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$.

Marginal distribution

- Given $p_{X,Y}$, how to find the distribution of each of the coordinates, that is p_X and p_Y?

Independence of r.v.'s

Definition

Discrete r.v.'s X, Y *are independent if for every* $x, y \in \mathbb{R}$ *the events* $\{X = x\}$ *a* $\{Y = y\}$ *are independent. That happens if and only if*

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

Product of independent r.v.'s

Theorem

For independent discrete r.v.'s X, Y we have

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Function of a random vector

Theorem

Suppose X, Y are r.v.'s on $(\Omega, \mathcal{F}, P)$, let $g : \mathbb{R}^2 \to \mathbb{R}$ be a function.

- *Then $Z = g(X, Y)$ is a r.v. on $(\Omega, \mathcal{F}, P)$*
- *and it satisfies*

$$\mathbb{E}(g(X, Y)) = \sum_{x \in ImX} \sum_{y \in ImY} g(x, y) P(X = x, Y = y),$$

whenever the sum is defined.

Theorem

For X, Y r.v.'s and $a, b \in \mathbb{R}$ we have

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Proof of the theorem about variance

Sum of independent r.v.'s

- Given $p_{X,Y}$, how to find the distribution of the sum, $Z = X + Y$?

Sum of independent r.v.'s – convolution

Theorem

Let X, Y be discrete random variables. Then their sum $Z = X + Y$ has PMF given by

$$P(Z = z) = \sum_{x \in Im(X)} P(X = x, Y = z - x).$$

If we further assume that X, Y are independent, then

$$P(Z = z) = \sum_{x \in Im(X)} P(X = x)P(Y = z - x).$$