2 Discrete Random Variables

2.1 Random Variables

Slide 1 "→" is an arrow symbol that indicates the mapping of outcomes in the sample space to real numbers by the random variable.

Random Variables

Chance experiments: Recall, we defined the sample space Ω = all possible (elementary) outcomes ω, and probability function $\Pr(E)$ for events $E \subseteq \Omega$.

Definition 2.1: Random variable (r.v.). a real-valued function $X(\omega)$ of outcomes, that is, a function

X(ω) input individual outcome from Ω, and use the function to map the ω to result(real numbers)

$$X : \Omega \to \Re.$$

Discrete r.v: If X takes only finintly or countably many values (often, the integers).

sample space -> real number

Example: rolling two dice, with, e.g., ω = (⚀, ⚂).
Define X = sum of the two dice. Then, $X(\omega) = 4$, etc.
Or Y = difference. Then $Y(\omega) = -2$, etc.

Events of special interest are then, e.g., $A = \{X = 4\}$, that is, the event of all possible outcomes ω that are mapped to $X(\omega) = 4$.

real number -> sample space

Example (ctd.): $\{X = 4\}$ = {(⚀⚂), (⚁⚁), (⚂⚀)}.

"X" represents a random variable. A random variable, denoted as "X," is a function that assigns real numbers to the outcomes in a sample space.

"ω" (omega) represents an individual outcome. In contrast, "ω" represents a specific outcome within the sample space of a random experiment.

2.2 Independent Random Variables

Slide 2

Independent R.V.'s

Use prob function $\Pr(A)$ with $A = \{X = x\}$ for a r.v. X.
It's important and common enough to give it a name :-)

Probability (mass) function (p.m.f.):

$$p_X(a) = \Pr(X = a)$$

(or just $p(a)$ when X is clear from the context).

(Cumulative) Distribution function: we often use

$$F_X(a) = \Pr(X \le a)$$

Note, for an integer-valued r.v., $p_X(a) = F_X(a) - F_X(a - 1)$.

Joint prob function: Similarly for a pair (X, Y) of (jointly distributed) r.v.'s we define

$$p_{X,Y}(a, b) = \Pr\left((X = a) \cap (Y = b)\right)$$

For short, we often write just $\Pr(X = a, Y = b)$.

Slide 3

Lemma 2.0: $p_X(a) = \sum_b p_{X,Y}(a, b)$ Law of Total Probability

Proof: this is just the law of total probability with $E_b = \{Y = b\}$.
The rest is just about notation:

$$p_X(a) = \Pr(X = a) = \sum_b \Pr\{(X = a) \cap (Y = b)\} = \sum_b p_{X,Y}(a, b)$$

Example: rolling two dice, X = sum of the faces,
$p_X(2)$ = Pr{(⚀, ⚀)} $= \frac{1}{36}$, $p_X(3) = \frac{2}{36}$ etc.
Let Y = difference of the faces,
$p_{X,Y}(2, 0) = p_{X,Y}(3, 1) = p_{X,Y}(3, -1) = \ldots = p_{X,Y}(12, 0) = \frac{1}{36}$.

Slide 4

Def 2.2: Independent r.v.'s. using the definition of independent events we say X and Y are independent iff

$$p_{X,Y}(x, y) = \Pr(X = x, Y = y) = \Pr(X = x) \cdot \Pr(Y = y) = p_X(x) \cdot p_Y(y)$$

for all values x and y, using independence of the events $A = \{X = x\}$ and $B = \{Y = y\}$.

Mutually independent r.v's: Similarly, $X_1, \ldots, X_k$ are *mutually independent* if for any $I \subseteq \{1, \ldots, k\}$ we have

I is subset

$$p_{X_1 \ldots X_k}(x_1, \ldots, x_k) = \Pr\left(\bigcap_{i \in I} X_i = x_i\right) = \prod_{i \in I} p_{X_i}(x_i)$$

if the joint probability = product of the marginal probabilities

Example: 2 dice with X_1 = first die, X_2 = 2nd die.

We call X1 and X2 are two independent r.v

$$p_{X_1,X_2}(1, 3) = \frac{1}{36} = p_{X_1}(1) \cdot p_{X_2}(3) = \frac{1}{6} \cdot \frac{1}{6}$$

When we use both the joint probability, the probability of one variable at a time only we often refer to the letter as a marginal probability. So pXi (xi) we also call marginal probability.

2.3 Expectations

Slide 5

Expectation

The concept of r.v.'s is so useful because they attach numeric values to chance experiments, which can then be manipulated and summarized as real numbers.

Def 2.3: Expectation. The expectation of a discrete r.v. X is the average value, averaging over all possible values x, weighted with the corresponding probability

$$E(X) = \sum_x x \cdot p_X(x).$$

If the sum does not converge, we say the expectation is "unbounded".

Derivative: f'(h) = lim h→0 [f(x+h)-f(x)] / h
or
f'(x) = dy/dx (change of y / change of x)
d = delta, meaning is change

Slide 6 f(x) = x^2, want to know f'(x)
dy/dx = lim h→0 [f(x+h)−f(x)] / h
= lim h→0 [(x+h)^2−x^2] / h
= lim h→0 [h(2x+h)] / h
= lim h→0 (2x+h)
= 2x+0 = 2x

Example 1: back to 2 dice, and X = sum of the two dice.

$$E(X) = \frac{1}{36} \cdot 2 + \ldots\ldots + \frac{2}{36} \cdot 11 + \frac{1}{36} \cdot 12 = 7$$

Example 2: Consider another r.v. Y with $\Pr(Y = 2^i) = 1/2^i$, $i = 1, 2, \ldots$. Find $E(Y)$?

$$EY = \sum_{i=1}^{\infty} 2^i \cdot \frac{1}{2^i} = \infty.$$

We say Y has no expectation.

2.4 Linearity of Expectations

Slide 7

Linearity of Expectation

Being defined as a sum, expectations inherit any properties of sums, e.g.

Lemma 2.2: for any constants a, b,

$$E(aX + b) = aE(X) + b.$$

This is easily proven by writing out the sum, factoring a and noting $\sum_i b\, p_X(i) = b \sum p_X(i) = b \cdot 1$. f(x) = 5, then E(X) = b*1 = 5

In other words, for a linear function $g(\cdot)$

$$E[g(x)] = g(EX)$$

Slide 8

Theorem 2.1: Linearity of Expectation. For r.v.'s $X_1, \ldots, X_n$, consider a (new, derived) r.v. $Y = \sum_{i=1}^{n} X_i$. Then

Expectation of the sum = Sum of the expectation

$$E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

Integral of the joint probability density = Marginal probability density.
(mutually disjoint events in Ω)

Proof: for $n = 2$ variables X, Y with prob function $p(i, j)$. Recall the law of total probability for r.v.'s, $\sum_j p_{X,Y}(i, j) = p_X(i)$.

2∑i 3∑j (i*j) = (1·1)+(1·2)+(1·3)+(2·1)+(2·2)+(2·3)

$$\begin{aligned} E(X+Y) &= \sum_i \sum_j (i + j) p(i,j) = \sum_i \sum_j i\, p(i,j) + \sum_j \sum_i j\, p(i,j) \\ &= \sum_i i \sum_j p(i,j) + \sum_j j \sum_i p(i,j) \\ &= \sum_i i\, p_X(i) + \sum_j j\, p_Y(j) = E(X) + E(Y) \end{aligned}$$

2 dice, X represent the result first die and Y represent the result of second die.
E(X+Y) = ∑i∑j (i+j)p(i,j) = 2·P(X=1, Y=1)+3·P(X=1,Y=2)+…+12·P(X=6,Y=6) = 7

= ∑i∑j i*p(i,j) + ∑i∑j j*p(i,j) = [1·P(X=1, Y=1)+1·P(X=1, Y=1)]
+[1·P(X=1,Y=2)+2·P(X=1,Y=2)]+…+[6·P(X=6,Y=6)+6·P(X=6,Y=6)] = 7

= ∑i [∑j p(i,j)] + ∑j [∑j p(i,j)] = 1·P(X=1, Y=1)+1·P(X=1, Y=2)+1·P(X=1, Y=3)+…
1·P(X=1, Y=1)+1·P(X=3, Y=1)+1·P(X=3, Y=1) = 7

2.5 Examples

Slide 9

Examples

For the following calculations we will need the identities

$$\sum_{j=1}^{k} j^2 = \frac{k(k+1)(2k+1)}{6} \text{ and } \sum_{1}^{k} j = \frac{k(k+1)}{2}.$$

 # 2.1: roll a fair k-sided die with the numbers 1 through k. Let X = number that appears. Find $E(X)$ =?

Solution: $E(X) = \sum_{j=1}^{k} j \frac{1}{k} = \frac{k+1}{2}$ 1+2+3+…k=k(k+1)/2

 # 2.9a: rolling the k-sided die twice, let X_1 and X_2 denote the number that appears. Find $E[\max(X_1, X_2)]$?

Solution: Let $M = \max(X_1, X_2)$. First find

$$F_M(j) = \Pr(X_1 \le j, X_2 \le j) = (j/k)^2$$

and therefore
$p_M(j) = F_M(j) - F_M(j-1) = \frac{j^2-(j-1)^2}{k^2} = \frac{2j-1}{k^2}$.

$$\Longrightarrow E(M) = \sum_{j=1}^{k} j \frac{2j-1}{k^2} = \frac{2}{k^2} \sum j^2 - \frac{1}{k^2} \sum j = \ldots$$

2.6 Jensen's Inequality

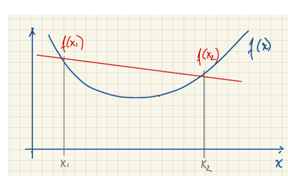
Slide 10

Jensen's Inequality

Definition 2.4: Convex functions. A function $f : \Re \to \Re$ is *convex* if for any x_1, x_2 and $0 \le \lambda \le 1$

$$f(\lambda x_1 + (1-\lambda) x_2) \quad \le \quad \lambda f(x_1) + (1-\lambda) f(x_2)$$

function betw x_1, x_2 line segment betw $(x_1, f(x_1)), (x_2, f(x_2))$



If f is twice differentiable, f is convex $\iff$ $f''(x) \ge 0$.

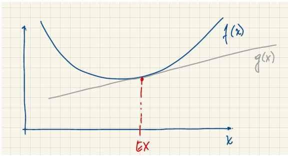
Theorem 2.4: Jensen's Inequality. If $f(\cdot)$ is a *convex* function, then

$$E[f(X)] \ge f(E[X])$$

Expected value on the curve ≥ F evaluated for the expected value of x.

Slide 11

Why is this true? Let $\bar{x} = EX$.



By Lemma 2.2,

$$E[g(x)] = g(\bar{x}) = f(\bar{x})$$

for the linear function $g(x)$.

We have

$$g(x) \le f(x) \implies E[f(x)] \ge E[g(x)] = f(\bar{x}).$$

(formal) Proof of Th 2.4: $\rightarrow$ book.
In short, use a Taylor series expansion, using the mean-value form of the remainder.

2.7 Binomial R.V's

Slide 12

Binomial R.V's

Bernoulli r.v.: a binary r.v. $Y \in \{0, 1\}$ with

$$p_Y(y) = \begin{cases} p & \text{for } y = 1 \text{ "success"} \\ (1-p) & \text{for } y = 0 \text{ "failure"} \end{cases}$$

We write $Y \sim \text{Bern}(p)$. Note: $E(Y) = p \cdot 1 + (1-p) \cdot 0 = p$.

Binomial experiments: Many experiments can be described as counting the number of successes ($Y_i = 1$) in a fixed number (n) of Bernoulli trials. For example,

- Flipping n coins, and counting $X = \#$ heads;
- Treating n patients, and recording $X = \#$ of patients who respond;
- Observing change in stock price over n days, and recording $X = \#$ days it rises; etc.

All these have a common structure, and we can argue for a prob function for X.

Slide 13

Binomial experiments: Binom r.v. X arise when we

- repeat a basic (Bernoulli) experiment with $Y_i \in \{0, 1\}$,
- independently, with always the same $p_{Y_i}(1) = p$ (success), i.e., $Y_i \sim \text{Bern}(p)$,
- a *fixed* number of times (n),
- and $X = \sum Y_i$ counts the number of successes.

We write $X \sim \text{Bin}(n, p)$.

X: rv for counting the number of successes.
n: number of experiments
p: probability of success

To find $p_X(j)$ note

j: number of successes

-
$$\Pr\{(\underbrace{1, \ldots, 1}_{[j \text{ times}]}, \underbrace{0, \ldots, 0}_{[n-j \text{ times}]})\} = p^j (1-p)^{n-j}$$
and same for any other sequence of j successes and $n - j$ failures.

- There are $\binom{n}{j}$ such sequences.

Slide 14

Prob Function & Expectation of a Bin r.v.

Definition 2.5: Binomial r.v. $X \sim \text{Bin}(n, p)$ if

$$p_X(j) = \binom{n}{j} p^j (1-p)^{n-j}$$

Result: if $X \sim \text{Bin}(n, p)$, then $E(X) = np$

Proof: By Theorem 2.1.,

$$E(X) = \sum_{i=1}^{n} E(Y_i) = n \cdot p.$$

2.8 Examples

a+ar+ar^2+ar^3
The sum of a geometric series
S=a/(1-r)
S is the sum of the series.
a is the first term of the series.
r is the common ratio.

Slide 15

Example

A family has n children with probability αp^n $n \ge 1$, where $\alpha \le (1-p)/p$.

1. Let $X = \#$ of children. i.e., $p_X(n) = \alpha p^n$ for $n \geq 1$.
What proportion of families has no children? That is, find $p_X(0)$.

Recall a geometric series $S = \sum_{n \ge 0} q^n = \frac{1}{1-q}$, for $0 < q < 1$.
Then use

$$\begin{aligned} p_x(0) &= 1 - \sum_{n \ge 1} \alpha p^n = \\ &= 1 - \alpha p \sum_{\ell \ge 0} p^\ell = 1 - \frac{\alpha p}{1-p}. \end{aligned}$$

1 - ap^1 - ap^2 - ap^3 - ap^n

Slide 16

2. If each child is equally likely to be a boy or a girl (independently of each other), what proportion of families consist of k boys (and any number of girls)? That is, letting $Y = \#$ boys, find $p_Y(k)$.

Solution: We will use:

1. Law of total probability,
$p_Y(k) = \sum_n \Pr(Y = k \mid E_n) \Pr(E_n)$, with $E_n = \{X = n\}$.

n: number of children in this family

2. Recall again $S = \sum_{n\geq 0} q^n = \frac{1}{1-q}$,

$\implies \frac{dS}{dq} = \sum_{i\geq 1} i\, q^{i-1} = (1-q)^{-2}$, and in general

$$\frac{d^\ell S}{dq^\ell} = \sum_{i\geq \ell} i(i-1)\cdots(i-\ell+1)\, q^{(i-\ell)} = \ell!\ (1-q)^{-(\ell+1)}.$$

We will this with $q = (p/2)$ and $\ell = k$.

Slide 17

- For the law of total prob use $\Pr(E_n) = p_X(n)$.
- Note that for given n, $X \sim \text{Bin}(n, 1/2)$. That is, $\Pr(Y = k \mid E_n) = \binom{n}{k}(1/2)^n$ for $n \geq k$ (and 0 for $n < k$).

We get for $k \geq 1$:

$$\begin{aligned}
\Pr(Y = k) &= \sum_{n\geq k} \Pr(Y = k \mid E_n) \cdot \Pr(E_n) = \sum_{n\geq k} \binom{n}{k}(1/2)^n \cdot \alpha p^n \\
&= \frac{\alpha}{k!}\left(\frac{p}{2}\right)^k \sum_{n\geq k} n(n-1)\cdots(n-k+1)\left(\frac{p}{2}\right)^{(n-k)} \\
&= \frac{\alpha}{k!}\left(\frac{p}{2}\right)^k (k!)\left(1-\frac{p}{2}\right)^{-(k+1)} = \alpha(p/2)^k\left(1-\frac{p}{2}\right)^{-(k+1)}
\end{aligned}$$

And $p_Y(0) = 1 - \sum_{k\geq 1} p_Y(k)$.

2.9 Conditional Distribution

Slide 18

Conditional Distribution

Recall the definition of conditional probabilities $\Pr(A \mid B)$.

We use conditional probabilities for $A = \{Y = y\}$ and $B = \{X = x\}$ to define a conditional distribution and expectations.

Definition: Conditional distribution. we call

$$p_{Y|X}(Y = y \mid X = x) =$$

$$= \Pr(Y = y \mid X = x) = \frac{\Pr(Y = y, X = x)}{\Pr(X = x)} =$$

Conditional Probability: Pr(E I F) = Pr(E ∩ F)/Pr(F)

$$= \frac{p_{Y,X}(y, x)}{p_X(x)}$$

the conditional distribution of Y given X.
Then $p_{Y|X}$ is a probability mass function.

"Conditional prob's are *probabilities*" $\implies$ all results for prob's apply.

Slide 19

Example

For r.v.'s X and Y let $p(x, y)$ denote the joint probability function, with

$$\begin{array}{ll} p(0,0) = .4 & p(0,1) = .2 \\ p(1,0) = .1 & p(1,1) = .3 \end{array}$$

Find $p_{X|Y}(x \mid Y = 1)$.

PY(a) = Pr(Y = a)

Solution: First find $p_Y(1) = .2 + .3 = .5$, giving

Joint probability: Pr(X = x, Y = 1) = P(x, 1)

$$p_{X|Y}(x \mid Y = 1) = \frac{p(x, 1)}{p_Y(1)} = \begin{cases} 2/5 = 0.4 & \text{for } x = 0 \\ 3/5 = 0.6 & \text{for } x = 1. \end{cases}$$

Pr(X = x I Y = 1)

2.10 Conditional Expectation

Slide 20

Conditional Expectation

Since cond probabilities are probabilities $\implies$ can define expectation, as before

Definition 2.6: Conditional expectation.

$$E(Y \mid Z = z) = \sum_y y\, p_{Y|Z}(y \mid z)$$

Note that $E(Y \mid Z = z)$ is a number $\in \Re$.

Example: rolling two dice: Y = number on 1st die, and X = sum of the numbers on both. Then

$$E(X \mid Y = 2) = \sum_x x \underbrace{\Pr(X = x \mid Y = 2)}_{1/6\ \ \Pr(X=x,Y=2)\,/\,PY(2)\ \ 1} = \sum_{x=3}^{8} x \cdot \frac{1}{6} = \frac{11}{2}$$

For later reference, recall that $E(X \mid Y = 2) = 5.50$ is a number. If we were not told $Y = 2$, we would just have

$$\sum_{x=Y+1}^{Y+6} \ldots = 3.50 + Y.$$

Slide 21

Lemma 2.5: the average conditional expectation = expectation,

$$E(X) = \sum_y E(X \mid Y = y)\, p_Y(y).$$

E(X)=∑ x·P(X=x)

Proof: Use the law of total prob with $E_y = \{Y = y\}$ to get

$$p_X(x) = \sum_y \Pr(X = x \mid Y = y)\, p_Y(y)$$

and therefore

$$\begin{aligned}
E(X) = \sum_x x\, p_X(x) &= \sum_x \left\{\sum_y x \Pr(X = x \mid Y = y)\, p_Y(y)\right\} \\
&= \sum_y \left\{p_Y(y) \sum_x x \Pr(X = x \mid Y = y)\right\} \\
&= \sum_y p_Y(y)\, E(X \mid Y = y)
\end{aligned}$$

Slide 22

Conditional expectation as a r.v.

Recall that a r.v. $Y : \Omega \to \Re$ is a real-valued function, and we use $Y = y$ to indicate a specific realization $y \in \Re$.

A second (and *different*) definition of *conditional expectation* is as a function of Y:

In $E(X \mid Y = y)$, if we remove the $= y$, we are left with a function of the r.v. Y:

One type of conditional expectation which is just a number

Example: Recall the 2 dice, Y = 1st die, and X = sum of the two.

$$E(X \mid Y = 2) \;=\; 5.5 \qquad\qquad E(X \mid Y \quad) \;=\; Y + \frac{7}{2}$$

$$5.5 \qquad\qquad \{⚀, ⚁, \cdots, ⚅\} \;\mapsto\; \Re$$

a value $\in \Re$ — **Type one**

a function $\Omega \mapsto \Re$ (a r.v.) — **Type two**

Slide 23

Second: mapping from whatever the first die shows to the real numbers

Definition 2.7: Conditional expectation. $E(X \mid Y)$ is a r.v., which takes value $E(X \mid Y = y)$ when $Y = y$.

i.e. $W = E(X \mid Y)$ is a r.v., not a value! It's a mapping $\Omega \to \Re$.

Theorem 2.7: $E(X) = E[E(X \mid Y)]$

Proof: $E(X \mid Y) = f(Y)$ with $f(Y) = f(y)$ when $Y = y$

The average of conditional expectations is the marginal expectation.

$$\begin{aligned}
\Longrightarrow\; E\,[E(X \mid Y)] &= E[f(Y)] = \\
&= \sum_y f(y)\, p_Y(y) \\
&= \sum_y E(X \mid Y = y)\, p_Y(y) = E(X).
\end{aligned}$$

The last equality is by Lemma 2.5.

2.11 Examples

Slide 24

Example

We repeatedly roll a fair die. Let X = # of rolls until the first ⚅ (including the ⚅ itself), and Y = # of rolls until the first ⚄.

(a) Find $p_X(x)$, $p_{X|Y}(x \mid Y = 1)$ and $p_{X|Y}(x \mid Y = 2)$.

E(Y | Z = z) = Σy pY|Z(y | z)

probability of rolling a six when it's not a five.

Solution: Let $\pi = 1/6$ and $q = 1 - \pi$, and $\tilde{\pi} = 1/5$. Then

probability of X = j

$$\begin{aligned}
p_X(j) &= (1-\pi)^{j-1}\pi, \\
p_{X|Y}(X = j \mid Y = 1) &= (1-\pi)^{j-2}\pi,\; j = 2, \ldots, \\
p_{X|Y}(X = j \mid Y = 2) &= \begin{cases} \tilde{\pi} & j = 1 \\ (1-\tilde{\pi})(1-\pi)^{j-3}\pi, & j \geq 3 \end{cases}
\end{aligned}$$

the probability of a non-six on the first one is 4/5,

$$\bullet, ⚄, \underbrace{\bullet, \cdots \bullet}_{(j-3)\times}, ⚅$$

PY(Y = 1) = 1/6, your conditioning on the first roll being a five.

Slide 25

a+ar+ar^2+ar^3
The sum of a geometric series
S=a/(1-r)
S is the sum of the series.
a is the first term of the series.
r is the common ratio.

(b) Find $E(X)$, $E(X \mid Y = 1)$ and $E(X \mid Y = 2)$.

Solution:

= π * d[1/(1-q)]/ dq

$$E(X) = \pi \sum_{j=1}^{\infty} j\,q^{\,j-1} = \pi \frac{d}{dq} \sum_{j=0}^{\infty} q^{\,j} = \frac{\pi}{(1-q)^2} = \frac{1}{\pi}$$

(see also later discussion on the geometric distribution). Next

q=1−π

$$E(X \mid Y = 1) = \pi \sum_{j=2}^{\infty} j\,q^{\,j-2} = \pi \sum_{\ell=1}^{\infty} (\ell+1)\,q^{\,\ell-1} =$$

use geometric series = 1

$$= \pi \sum_{\ell=1}^{\infty} \ell\,q^{\,\ell-1} + \pi \sum_{\ell=1}^{\infty} q^{\,\ell-1} = EX + 1$$

For the last equality use $\pi q^{\ell-1} = p_X(\ell)$. Similarly,

$$E(X \mid Y = 2) = \tilde{\pi} \cdot 1 + (1-\tilde{\pi}) \sum_{j=3}^{\infty} j\,q^{\,j-3}\pi =$$

$$\tilde{\pi} + (1-\tilde{\pi}) \sum_{\ell=1}^{\infty} (\ell+2)q^{\,\ell-1}\pi = \tilde{\pi} + (1-\tilde{\pi})(EX+2)$$

2.12 Example: Recursive Function Calls

Slide 26

X ~ Bin(n, p)
X: number of successes
n: number of experiments
p: probability of success

Example 4: Recursive function calls

Setup: a function includes Y_1 recursive calls to itself. If $Y_1 \sim \text{Bin}(n, p)$, find the expected total # of calls to the function?

Calls in generation i: Let Y_i = number of calls in generation i (spawned by another call in generation $i-1$).

Let X_k = # of calls spawned by the k-th call in the $(i-1)$-st generation, $k = 1, \ldots, Y_{i-1}$.
Then also $X_k \sim \text{Bin}(n, p)$, and therefore

$$E(Y_i \mid Y_{i-1} = y_{i-1}) = \sum_{k=1}^{y_{i-1}} E(X_k \mid Y_{i-1} = y_{i-1}) = \sum_k np = y_{i-1}{\cdot}np$$

harmonic number: sum of recipr

Slide 27

Example ④: Recursive function calls

Expectation $E(Y_i)$: Using Theorem 2.7 we get

$$E(Y_i) = E\{E(Y_i \mid Y_{i-1})\} = E(Y_{i-1} np) = np\, E(Y_{i-1})$$

and therefore, starting with $Y_0 = 1$, by induction $E(Y_i) = (np)^i$.

Total # calls:

$$E\left(\sum_{i=1}^{\infty} Y_i\right) = \sum_{i=1}^{\infty} E(Y_i) = \sum_i (np)^i.$$

If $np < 1$, the expected total # calls converges; otherwise it diverges (and our program crashes ...).

2.13 Geometric Distribution

Slide 28

Geometric Distribution

An important detail about binomial experiments is the fixed number of repetitions (of the binary experiment). This is violated, for example, if we flip a coin until we see a head.

Definition 2.8: Goemetric r.v. A geometric r.v. with parameter p is defined by the probability distribution

X: the x times to do the experiments

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

It arises, for example, if we count the # coin flips until the first head.

Lemma 2.8: Memoryless property.
$\Pr(X = n + k \mid X > k) = \Pr(X = n)$

Proof: exercise. Use $\Pr(X > k) = (1-p)^k$ (need k failures for $X > k$).

For later reference, $\Pr(X \geq k) = (1-p)^{k-1}$

Slide 29

Mean of Geom(p)

Lemma 2.9: If $X > 0$ is a discrete r.v., then

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

Proof:

$$\sum_{i=1}^{\infty} \Pr(X \geq i) = \sum_{i=1}^{\infty}\left(\sum_{j=i}^{\infty} \Pr(X = j)\right) = \sum_{j=1}^{\infty}\sum_{i=1}^{j} \Pr(X = j) = \sum_j j \Pr(X = j) = E(X)$$

$E(X)$: expectation of a Geom(p) r.v. X:

$$E(X) = \sum_{i=1}^{\infty} \Pr(X \geq i) = \sum_{i=1}^{\infty} (1-p)^{i-1} = \boxed{\frac{1}{1-(1-p)}} = 1/p.$$

2.14 Example: Coupon Collector's Problem

Slide 30

Example ⑤: Coupon Collector's Problem

The following setup is encountered in many problems. The coupon collector is just the traditional story around it.

Setup: Assume each box of cereal includes a coupon, randomly chosen from n possible coupons.
How many boxes to get a complete set of all n coupons?

Let X_i = number of boxes bought while you had $i-1$ coupons. Then $X = \sum_{i=1}^{n} X_i$ is the total number of boxes you need.

Geometric r.v.: $X_i \sim \text{Geom}(p_i)$, $p_i = \frac{n-(i-1)}{n}$ p1 = 1, p2 = (n-1)/n
(since you already have $i-1$ coupons), and therefore $E(X_i) = 1/p_i = n/(n-i+1)$.

i = 1, then[n-i+1] = n
i = n, then[n-i+1] = 1

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \underbrace{\sum_{i=1}^{n} \frac{1}{i}}_{H(n)}.$$ n*1/1+n*1/2+...

X_1 is the number of boxes that we have to buy for the first coupon. X_1 = 1
X_3 is the number of boxes that you have to buy until you get third distinct coupon

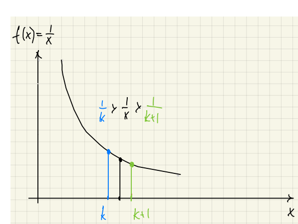
Slide 31

harmonic number: sum of reciprocals from 1 to n

Coupon Collector (ctd.)

Lemma 2.10: $H(n) = \ln n + \Theta(1)$

Proof: noting $\frac{1}{k+1} \leq \frac{1}{x} \leq \frac{1}{k}$ for $k \leq x \leq k+1$,



d/dx* (ln(x))= 1/x

$$\ln n = \int_{x=1}^{n} \frac{1}{x} dx \leq \sum_{k=1}^{n-1} \frac{1}{k} < \sum_{k=1}^{n} \frac{1}{k} = 1 + \sum_{k=2}^{n} \frac{1}{k} \leq 1 + \int_{x=1}^{n} \frac{1}{x} = 1 + \ln n$$

$\Longrightarrow$ $\ln n \leq H(n) \leq \ln n + 1.$ H(n)

Split the LHS(∫(from x = 1 to n) (1/x) dx) integral in a sum of integrals from 1 to 2, 2 to 3, ..., n - 1 to n. On 1 to 2, the function 1/x is <= 1/1, on 2 to 3, 1/x <= 1/2 ... Hence the sum of integrals <= RHS(∑(from k = 1 to n-1)1/k)

2.15 Example: Expected Run Time of Quicksort

Slide 32 Big O (O()): upper bounds
Big Omega (Ω()): lower bounds

Example 6: Expected Run Time of Quicksort

Quicksort: recursively sort a list $S = \{x_1, \ldots, x_n\}$:

- If $|S| \le 1$, return S, otherwise
- Randomly select $y \in S$ ("pivot"), let $S_1 = \{x \in S : x < y\}$ and $S_2 = \{x \in S : x \ge y\}$.
- Return quicks(S_1) ∪ $\{y\}$ ∪ quicks(S_2).

Claim: with random y, # of comparisons is

$$2n \ln n + O(n).$$

At worst Ω(n^2) if y = min(S)(or max(S)),
At best O(n log n) if y = median(S) (w/o proof)

Slide 33

Proof

First some observations:

- Let $y_1, \ldots, y_n$ denote the ordered elements $x_1, \ldots, x_n$; let $X_{ij} = \begin{cases} 1 & \text{if } x_i, x_j \text{ are compared} \\ 0 & \text{otherwise} \end{cases}$ and similarly Y_{ij}.
- Total # comparisons $X = \sum_{i<j} X_{ij} = \sum Y_{ij}$.
- Note $EX_{ij} = \Pr(X_{ij} = 1)$, and same for Y_{ij}.
- $Y_{ij} = 1 \iff y_i$ or y_j is the first pivot selected from $\underbrace{\{y_i, \ldots, y_j\}}_{j-i+1 \text{ \#'s}}$

  $\implies \Pr(Y_{ij} = 1) = \frac{2}{j-i+1}$

Slide 34

We have

$$\begin{aligned} E(X) &= \sum_{i=1}^{n-1} \sum_{j>i} EX_{ij} = \sum_{i=1}^{n-1} \sum_{j>i} EY_{ij} = \\ &= \sum_{i=1}^{n-1} \sum_{j>i} \frac{2}{\underbrace{j-i+1}_{k}} = \sum_{k=2}^{n} \sum_{i=1}^{n+1-k} \frac{2}{k} = \sum_{k=2}^{n} (n+1-k)\frac{2}{k} \\ &= 2(n+1)\sum_{k=2}^{n} \frac{1}{k} - 2(n-1) = (2n+2)\underbrace{\sum_{k=1}^{n} \frac{1}{k}}_{H(n)} - 2(n-1) \end{aligned}$$

Recall $\sum_{k=1}^{n} \frac{1}{k} = H(n) = \ln n + \Theta(1) \implies EX = 2n \ln n + \Theta(n)$.

2.16 Poisson distribution

Slide 35

Poisson Distribution

Note: this is §5.3 in the book.

Definition 5.1: Poisson distribution. A discrete r.v. X with

$$p_X(j) = \frac{e^{-\lambda}\lambda^j}{j!}, \quad \text{for } j = 0, 1, 2, \ldots$$

We write $X \sim \text{Poi}(\lambda)$.

General setup: Poisson probabilities are good approximations for probabilities in many problems. For example:

Poisson probabilties can be used as an approximation for Binomial probabilities with large n and small p (such that np remains moderate):

If $X \sim \text{Bin}(n, p)$ with $np = \lambda$ (or $p = \lambda/n$) and large n, then $p_X(i) \approx e^{-\lambda}\frac{\lambda^i}{i!}$, i.e., approximated by Poi(λ) probabilities.

Proof: exercise, problem 7[1]

"X ~ Poi(λ)" means that the random variable X follows a Poisson distribution with a mean rate of λ

2.17 Examples

Slide 36

Examples

1. The probability of a "three of a kind" in poker is approximately $p = 1/50$.

Use the Poisson approximation to estimate the probability you will get at least one "three of a kind" if you play $n = 20$ hands.

That is, letting X = # hands with three of a kind, find $\Pr(X \ge 1)$. Get 0 "three of a kind" if play n = 20 hands. So j=0

Solution: use the Poi approx with $\lambda = np = 2/5$ to get

$$\Pr(X \ge 1) = 1 - p_X(0) \approx 1 - e^{-\lambda} = 1 - e^{-0.4} = 0.33.$$

X ~ Bin(n, p) = X ~ Bin(20, 1/50) = pX(0) = C(0 50)(1/50)^0(49/50)^20
= 0.66760797175
.

Slide 37

2. Let N_t = # of earthquakes in the western portion of the United States in t weeks. Assume $N_t \sim \text{Poi}(\lambda t)$, with $\lambda = 2$. That is earthquakes occur at a rate of $\lambda = 2$ per week.

(a) Find the probability of ≥ 3 earthquakes in 2 weeks.

Solution: Let $X = N_2 \sim \text{Poi}(4)$. Then

$$\Pr(X \ge 3) = 1 - p_X(0) - p_X(1) - p_X(2) = \ldots = 1 - 13e^{-4}.$$

[1] problem # reference in the video is outdated

pX(0) = e^(-4)
pX(1) = e^(-4)λ
pX(2) = [e^(-4)λ^2]/2

(b) Letting Y = time until next earthquake, find $\Pr(Y \le t)$.

Hint: Find first $\Pr(Y > t) = \Pr(N_t = 0) = \ldots$.

Solution: Use $N_t \sim \text{Poi}(\lambda t)$ to get

$$\Pr(Y > t) = p_{N_t}(0) = e^{-\lambda t} \implies \Pr(Y \le t) = 1 - e^{-\lambda t}.$$

Note: Did you notice that the r.v. Y here is not a discrete r.v? We will talk more about this in Chapter 8.

2.18 More Properties of Poi R.V's

Slide 38

More Properties of Poi R.V's

Problem 2. is an example for a (very common) application of Poisson r.v.'s to represent the number of certain events:

(1) For a *short time interval h* the probability of observing an event is proportional to the length of the interval, i.e. $\approx \lambda h$;
(2) prob of ≥ 2 events in a short interval $\to 0$ as $h \to 0$; and
(3) # of events in non-overlapping intervals are independent.

Let N_t = # events in a time interval t and assume

(1) $\Pr(N_h = 1) = \lambda h + o(h)$;

(2) $\Pr(N_h \ge 2) = o(h)$;

The "o(h)" term indicates that the probability of observing exactly one event becomes vanishingly small faster than "h" as "h" approaches zero.

(3) For any n non-overlapping time intervals the numbers of events in those intervals are independent.

Under these assumptions $N_t \sim \text{Poi}(\lambda t)$.

"Pr(Nh = 1)": the probability of exactly one event in the time interval "h."

"λh": This term is proportional to the length of the time interval "h" and represents the probability that one event occurs in the interval.

"o(h)": The probability of observing one event may deviate slightly from strict proportionality as the interval becomes infinitesimally small.

Slide 39

Sums of Poi r.v.'s

Lemma 5.2: Sums of Poi r.v.'s. If $X_i \sim \text{Poi}(\mu_i)$, $i = 1, \ldots, n$, independently, then $S = \sum_{i=1}^{n} X_i \sim \text{Poi}(\sum_{i=1}^{n} \mu_i)$.

Proof: For $n = 2 \to$ problem 9 [2]. By induction we get the desired result for $n > 2$.

[2] problem # given in the video is outdated