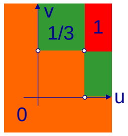
Discrete Random Variables

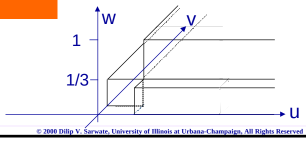
- Suppose that the discrete random variable **X** takes on values $u_1, u_2, \ldots, u_n, \ldots$ and the discrete random variable **Y** takes on values $v_1, v_2, \ldots, v_m, \ldots$
- (**X**, **Y**), the random point in the plane, takes on values (u_i, v_j)
- The joint CDF $F_{X,Y}(u,v)$ is a staircase function that has a step at each of the values (u_i, v_j) that (**X**, **Y**) takes on

An example of the CDF of (X, Y)

- Example: **X** and **Y** are Bernoulli random variables. However, P{(**X**,**Y**) = (0,0)} = 0



and the other three possible values (1,0), (0,1), (1,1) for (**X**,**Y**) have probability 1/3 each

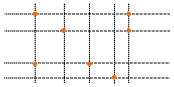


The joint probability mass function

- The joint probability mass function (joint pmf) for discrete random variables **X** and **Y** taking on values $u_1, u_2, \ldots, u_n, \ldots$ and $v_1, v_2, \ldots, v_m, \ldots$ respectively is defined as
  $p_{X,Y}(u,v) = P\{X = u_i, Y = v_j\}$ if $u = u_i$, $v = v_j$,
  and
  $p_{X,Y}(u,v) = 0$ otherwise
- The joint pmf describes a collection of point masses in the plane

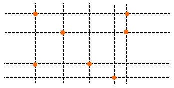
Some thoughts about the joint pmf

- The joint pmf defines a collection of point masses in the plane
- The point masses are at the intersections of the lines in the plane whose equations are $u = u_i$ and $v = v_j$
- The point masses lie on a grid



More thoughts about the joint pmf

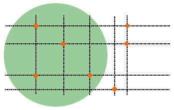
- The point masses lie on a grid



- Not every grid point need have a mass
- Total probability mass = 1
- Hence, $\sum_i \sum_j p_{X,Y}(u_i, v_j) = 1$
- $p_{X,Y}(u, v) \geq 0$ for all u and v, $-\infty < u,v < \infty$

Probabilities from the joint pmf

- Let A denote a region of the plane
- Then, P{(**X**, **Y**)   A} is the sum of all the probability masses in the region A

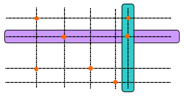


- This also holds if A is a curve (including a straight line as a special case) — just sum up all the probability masses on the curve

The marginal pmfs of X and Y

- $p_{\mathbf{X}}(u)$ and $p_{\mathbf{Y}}(v)$, the marginal pmfs of **X** and **Y**, are easily obtained from the joint pmf $p_{\mathbf{X},\mathbf{Y}}(u, v)$
- $p_{\mathbf{X}}(u_i) = \sum_j p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)$; $p_{\mathbf{Y}}(v_j) = \sum_i p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)$
- As with CDFs, the word marginal is not pejorative
- One possible reason for this unusual nomenclature will be presented real soon now (RSN)

Marginal pmfs from the joint pmf



- $p_{\mathbf{X}}(u_i) = \sum_j p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)$; $p_{\mathbf{Y}}(v_j) = \sum_i p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)$
- $p_{\mathbf{X}}(u_i) =$ sum of all probability masses lying on the vertical line with equation $u = u_i$
- $p_{\mathbf{Y}}(v_j) =$ sum of all probability masses lying on the horizontal line with equation $v = v_j$

The joint probability matrix

- Because of the grid structure of the joint pmf, it is convenient to think of the masses as entries in a matrix or array
  - Rows are labeled with the v_j's and columns are labeled with the u_i's
  - The matrix entry in the v_j-th row and u_i-th column is the probability of the event $P\{\mathbf{X} = u_i, \mathbf{Y} = v_j\}$, i.e. the value of $p_{\mathbf{X},\mathbf{Y}}(u_i, v_j)$

A picture is worth a thousand words

$p_{\mathbf{X},\mathbf{Y}}(u_i, v_j)$ is written as $p(u_i, v_j)$ for brevity

v_m	$p(u_1,v_m)$	$p(u_2,v_m)$	• • •	$p(u_n,v_m)$
•	•	•	• • •	•
•	•	•	• • •	•
•	•	•	• • •	•
•	•	•	• • •	•
v_3	$p(u_1,v_3)$	$p(u_2,v_3)$	• • •	$p(u_n,v_3)$
v_2	$p(u_1,v_2)$	$p(u_2,v_2)$	• • •	$p(u_n,v_2)$
v_1	$p(u_1,v_1)$	$p(u_2,v_1)$	• • •	$p(u_n,v_1)$
	u_1	u_2	• • •	u_n

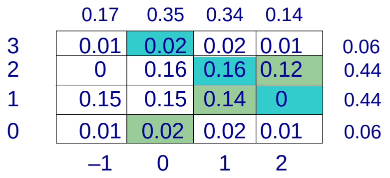
- Note the ordering of the v_j's

Why marginal, for crying out loud?

	$p_{\mathbf{X}}(u_1)$	$p_{\mathbf{X}}(u_2)$	• • •	$p_{\mathbf{X}}(u_n)$	
v_m	$p(u_1,v_m)$	$p(u_2,v_m)$	• • •	$p(u_n,v_m)$	$p_{\mathbf{Y}}(v_m)$
•	•	•	• • •	•	•
•	•	•	• • •	•	•
•	•	•	• • •	•	•
•	•	•	• • •	•	•
v_3	$p(u_1,v_3)$	$p(u_2,v_3)$	• • •	$p(u_n,v_3)$	$p_{\mathbf{Y}}(v_3)$
v_2	$p(u_1,v_2)$	$p(u_2,v_2)$	• • •	$p(u_n,v_2)$	$p_{\mathbf{Y}}(v_2)$
v_1	$p(u_1,v_1)$	$p(u_2,v_1)$	• • •	$p(u_n,v_1)$	$p_{\mathbf{Y}}(v_1)$
	u_1	u_2	• • •	u_n	

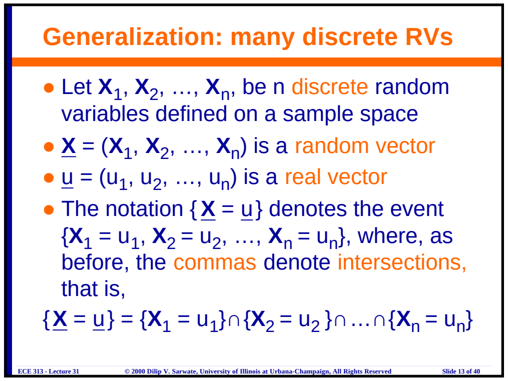
- Row and columns sum appear in the margins

An example



	0.17	0.35	0.34	0.14	
3	0.01	0.02	0.02	0.01	0.06
2	0	0.16	0.16	0.12	0.44
1	0.15	0.15	0.14	0	0.44
0	0.01	0.02	0.02	0.01	0.06
	–1	0	1	2	

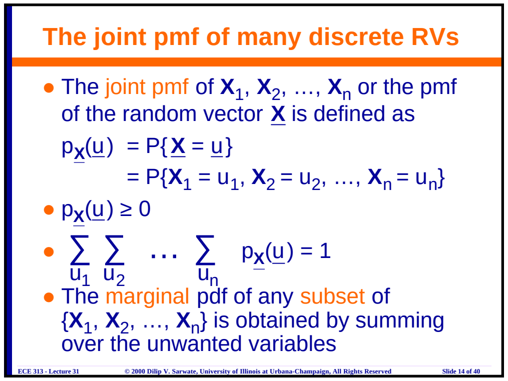
- The marginal pmfs are as shown
- $P\{\mathbf{X} = \mathbf{Y}\} = 0.02 + 0.14 + 0.12 = 0.28$
- $P\{\mathbf{X} + \mathbf{Y} = 3\} = 0.02 + 0.16 + 0 = 0.18$



Generalization: many discrete RVs

- Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, be n discrete random variables defined on a sample space
- $\underline{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ is a random vector
- $\underline{u} = (u_1, u_2, \ldots, u_n)$ is a real vector
- The notation $\{\underline{\mathbf{X}} = \underline{u}\}$ denotes the event $\{\mathbf{X}_1 = u_1, \mathbf{X}_2 = u_2, \ldots, \mathbf{X}_n = u_n\}$, where, as before, the commas denote intersections, that is,

$$\{\underline{\mathbf{X}} = \underline{u}\} = \{\mathbf{X}_1 = u_1\} \cap \{\mathbf{X}_2 = u_2\} \cap \ldots \cap \{\mathbf{X}_n = u_n\}$$

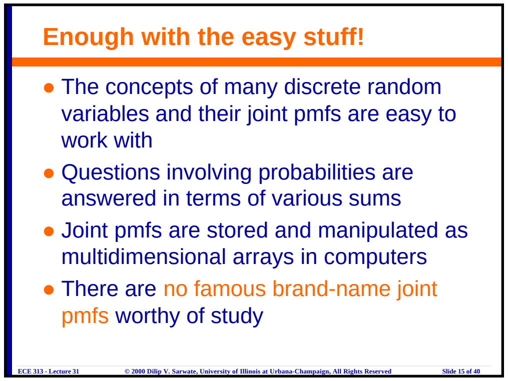


The joint pmf of many discrete RVs

- The joint pmf of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ or the pmf of the random vector $\underline{\mathbf{X}}$ is defined as

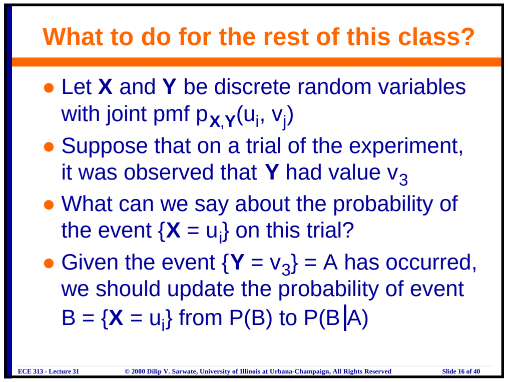
$$\begin{aligned} p_{\underline{\mathbf{X}}}(\underline{u}) &= P\{\underline{\mathbf{X}} = \underline{u}\} \\ &= P\{\mathbf{X}_1 = u_1, \mathbf{X}_2 = u_2, \ldots, \mathbf{X}_n = u_n\} \end{aligned}$$

- $p_{\underline{\mathbf{X}}}(\underline{u}) \geq 0$
- $\sum_{u_1} \sum_{u_2} \cdots \sum_{u_n} p_{\underline{\mathbf{X}}}(\underline{u}) = 1$
- The marginal pdf of any subset of $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$ is obtained by summing over the unwanted variables



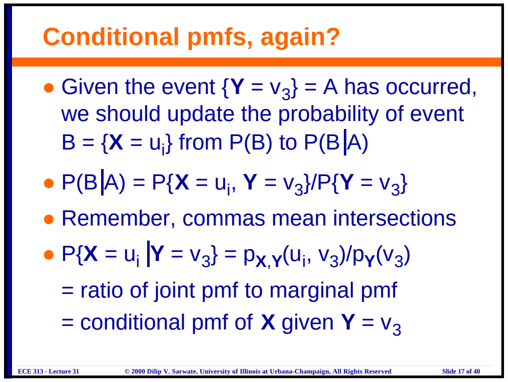
Enough with the easy stuff!

- The concepts of many discrete random variables and their joint pmfs are easy to work with
- Questions involving probabilities are answered in terms of various sums
- Joint pmfs are stored and manipulated as multidimensional arrays in computers
- There are no famous brand-name joint pmfs worthy of study



What to do for the rest of this class?

- Let $\mathbf{X}$ and $\mathbf{Y}$ be discrete random variables with joint pmf $p_{\mathbf{X},\mathbf{Y}}(u_i, v_j)$
- Suppose that on a trial of the experiment, it was observed that $\mathbf{Y}$ had value v_3
- What can we say about the probability of the event $\{\mathbf{X} = u_i\}$ on this trial?
- Given the event $\{\mathbf{Y} = v_3\} = A$ has occurred, we should update the probability of event $B = \{\mathbf{X} = u_i\}$ from $P(B)$ to $P(B|A)$

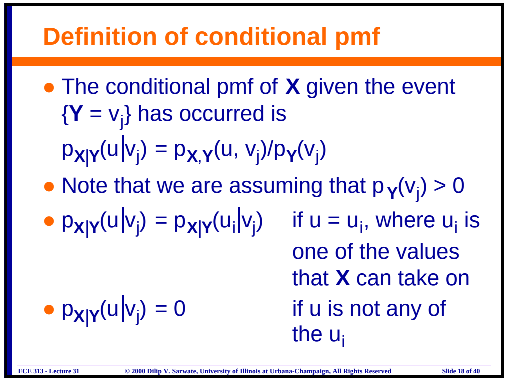


Conditional pmfs, again?

- Given the event $\{\mathbf{Y} = v_3\} = A$ has occurred, we should update the probability of event $B = \{\mathbf{X} = u_i\}$ from $P(B)$ to $P(B|A)$
- $P(B|A) = P\{\mathbf{X} = u_i, \mathbf{Y} = v_3\}/P\{\mathbf{Y} = v_3\}$
- Remember, commas mean intersections
- $P\{\mathbf{X} = u_i \,|\, \mathbf{Y} = v_3\} = p_{\mathbf{X},\mathbf{Y}}(u_i, v_3)/p_{\mathbf{Y}}(v_3)$

  = ratio of joint pmf to marginal pmf

  = conditional pmf of $\mathbf{X}$ given $\mathbf{Y} = v_3$



Definition of conditional pmf

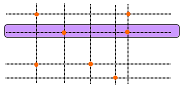
- The conditional pmf of $\mathbf{X}$ given the event $\{\mathbf{Y} = v_j\}$ has occurred is

  $p_{\mathbf{X}|\mathbf{Y}}(u|v_j) = p_{\mathbf{X},\mathbf{Y}}(u, v_j)/p_{\mathbf{Y}}(v_j)$
- Note that we are assuming that $p_{\mathbf{Y}}(v_j) > 0$
- $p_{\mathbf{X}|\mathbf{Y}}(u|v_j) = p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j)$ if $u = u_i$, where u_i is one of the values that $\mathbf{X}$ can take on
- $p_{\mathbf{X}|\mathbf{Y}}(u|v_j) = 0$ if u is not any of the u_i

Conditional pmf is a function of u !!

- The conditional pmf of **X** given the event $\{\mathbf{Y} = v_j\}$ has occurred is
  $p_{\mathbf{X}|\mathbf{Y}}(u|v_j) = p_{\mathbf{X},\mathbf{Y}}(u, v_j)/p_{\mathbf{Y}}(v_j)$
- The argument of the conditional pmf is u
- v_j is just the value of **Y** that was observed on this trial
- $p_{\mathbf{Y}}(v_j)$ is the value of the pmf of **Y** at v_j
- $p_{\mathbf{Y}}(v_j)$ is a number, not a function

Conditional pmf from the joint pmf



- $p_{\mathbf{X}|\mathbf{Y}}(u|v_j) = p_{\mathbf{X},\mathbf{Y}}(u, v_j)/p_{\mathbf{Y}}(v_j)$
- $p_{\mathbf{Y}}(v_j) =$ sum of all probability masses lying on the horizontal line with equation $v = v_j$
- $p_{\mathbf{X}|\mathbf{Y}}(u|v_j) = p_{\mathbf{X},\mathbf{Y}}(u, v_j)/p_{\mathbf{Y}}(v_j)$
  = ratio of the probability masses on the line to total mass on the line

Conditional pmfs are valid pmfs

- The conditional pmf of **X** given the event $\{\mathbf{Y} = v_j\}$ has occurred is
  $p_{\mathbf{X}|\mathbf{Y}}(u|v_j) = p_{\mathbf{X},\mathbf{Y}}(u, v_j)/p_{\mathbf{Y}}(v_j)$
- $p_{\mathbf{X}|\mathbf{Y}}(u|v_j) \geq 0$ for all values of u
- $\sum_i p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j) = 1$
- Note that v_j is just a constant in the sum, and that sum is 1 for all choices of v_j

A numerical example

	0.17	0.35	0.34	0.14	
3	0.01	0.02	0.02	0.01	0.06
2	0	0.16	0.16	0.12	0.44
1	0.15	0.15	0.14	0	0.44
0	0.01	0.02	0.02	0.01	0.06
	–1	0	1	2	

- $p_{\mathbf{X}|\mathbf{Y}}(u|2)$ has values 4/11, 4/11 and 3/11 at u = 0, 1, and 2; $p_{\mathbf{X}|\mathbf{Y}}(-1|2) = 0$
- $p_{\mathbf{X}|\mathbf{Y}}(u|3)$ has values 1/6, 1/3, 1/3 and 1/6 at u = –1, 0, 1, and 2

A sense of déjà vu all over again…

- We have already studied the notion of conditional pmf of **X** given an event A in Lecture 14
- Here, we are giving the conditioning event in terms of another random variable **Y**
- All the results from Lecture 14 hold — just write them in terms of **Y**
- Events $\{\mathbf{Y} = v_1\}, \{\mathbf{Y} = v_2\}, \ldots, \{\mathbf{Y} = v_m\}, \ldots$ are a partition of the sample space Ω

Unconditional pmf from conditional

- Events $\{\mathbf{Y} = v_1\}, \{\mathbf{Y} = v_2\}, \ldots, \{\mathbf{Y} = v_m\}, \ldots$ are a partition of the sample space Ω
- $p_{\mathbf{X}}(u_i) = \sum_j p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)$
  $= \sum_j p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j) \bullet p_{\mathbf{Y}}(v_j)$
- This is just the theorem of total probability
  $P(B) = \sum P(B|A_j)P(A_j)$
  expressed in terms of pmfs

Reversing the conditioning…

- $p_{\mathbf{X}}(u_i) = \sum_j p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)$

  $= \sum_j p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j)\bullet p_{\mathbf{Y}}(v_j)$
- $p_{\mathbf{Y}|\mathbf{X}}(v_j|u_i) = p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j)\bullet p_{\mathbf{Y}}(v_j)/\ p_{\mathbf{X}}(u_i)$
- This is just Bayes' formula

  $P(A|B) = P(B|A)P(A)/P(B)$

  expressed in terms of pmfs

Example

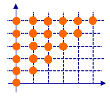
- **Y** takes on integer values 0, 1, 2, …
- The conditional pmf of **X** given that **Y** = n is a binomial pmf with parameters (n, p) where $0 < p < 1$
- Thus, conditioned on the event {**Y** = n}, the random variable **X** takes on the n+1 values 0, 1, 2, … n
- $p_{\mathbf{X}|\mathbf{Y}}(k|n) = \binom{n}{k}p^k(1-p)^{n-k}$, $0 \le k \le n$

Example (continued)

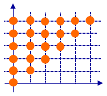
- $p_{\mathbf{X}|\mathbf{Y}}(k|n) = \binom{n}{k}p^k(1-p)^{n-k}$, $0 \le k \le n$
- Now suppose that **Y** is a Poisson random variable with parameter λ
- The joint pmf of **X** and **Y** is thus $p_{\mathbf{X},\mathbf{Y}}(k,n)$

$= p_{\mathbf{X}|\mathbf{Y}}(k|n)\bullet p_{\mathbf{Y}}(n) = \binom{n}{k}p^k(1-p)^{n-k}\bullet\exp(-\lambda)\bullet\frac{\lambda^n}{n!}$

$0 \le k \le n$



The joint pmf of X and Y

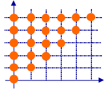


- For $0 \le k \le n$,

  $p_{\mathbf{X},\mathbf{Y}}(k,n) = p_{\mathbf{X}|\mathbf{Y}}(k|n)\bullet p_{\mathbf{Y}}(n)$

  $= \binom{n}{k}p^k(1-p)^{n-k}\bullet\exp(-\lambda)\bullet\frac{\lambda^n}{n!}$
- The point masses lie in the triangular region shown

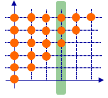
Some thoughts on the joint pmf



- $p_{\mathbf{X},\mathbf{Y}}(k,n) = \binom{n}{k}p^k(1-p)^{n-k}\bullet\exp(-\lambda)\bullet\frac{\lambda^n}{n!}$

  $0 \le k \le n$
- Conditioned on **Y** = n, **X** has values from 0 to n. However, unconditionally, **X** takes on values 0, 1, 2, … the same as **Y** !!
- **Y** ≥ **X** always

Unconditional (marginal) pmf of X



- $p_{\mathbf{X},\mathbf{Y}}(k,n) = \binom{n}{k}p^k(1-p)^{n-k}\bullet\exp(-\lambda)\bullet\frac{\lambda^n}{n!}$

  $0 \le k \le n$
- $p_{\mathbf{X}}(k) = \sum p_{\mathbf{X},\mathbf{Y}}(k,n)$ where sum is over n from n = k to ∞
- $p_{\mathbf{X}}(k) = \exp(-\lambda p)\bullet(\lambda p)^k/k!$ for k = 0, 1, 2, …
- **X** is a Poisson RV with parameter λp !!

Some details of the calculations

- $p_{\mathbf{X}}(k) = \sum p_{\mathbf{X},\mathbf{Y}}(k,n)$

$= \sum[\cancel{n!}/k!(n–k)!]•p^k(1–p)^{n–k}•\exp(–\lambda)•(\lambda)^n/\cancel{n!}$

$= \sum(\lambda p)^k(\lambda(1–p))^{n–k}•\exp(–\lambda)/k!(n–k)!$

- Remember that the sum is over n from n = k to ∞
- $p_{\mathbf{X}}(k) = \exp(–\lambda)•(\lambda p)^k/k!\sum(\lambda(1–p))^{n–k}/(n–k)!$

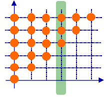
$= [\exp(–\lambda)•(\lambda p)^k/k!]•\exp(\lambda(1–p))$

$= \exp(–\lambda p)•(\lambda p)^k/k!$ for $k = 0, 1, 2, \ldots$

What have we learned so far?

- **Y** is a Poisson random variable with parameter λ
- Conditioned on **Y** = n, **X** is a binomial random variable with parameters (n,p)
- The joint pmf of **X** and **Y** is nonzero on a triangular region
- **Y** ≥ **X** always
- The unconditional pmf of **X** is Poisson with parameter λp

Conditional pmf of Y given X



- $p_{\mathbf{X},\mathbf{Y}}(k,n) = \binom{n}{k} p^k(1–p)^{n–k}•\exp(–\lambda)•\dfrac{\lambda^n}{n!}$
  $0 \le k \le n$
- $p_{\mathbf{Y}|\mathbf{X}}(n|k) = p_{\mathbf{X},\mathbf{Y}}(k,n)/p_{\mathbf{X}}(k)$

$= p_{\mathbf{X},\mathbf{Y}}(k,n)/\exp(–\lambda p)•(\lambda p)^k/k!$ for $n \ge k$

$= \exp(–\lambda(1–p))•(\lambda(1–p))^{n–k}/(n–k)!$ for $n \ge k$

- This is called a displaced Poisson pmf

Move to the right by k places…

- For $n \ge k$, $p_{\mathbf{Y}|\mathbf{X}}(n|k) = p_{\mathbf{X},\mathbf{Y}}(k,n)/p_{\mathbf{X}}(k)$

$= \exp(–\lambda(1–p))•(\lambda(1–p))^{n–k}/(n–k)!$

- This is a displaced Poisson pmf with parameter λ(1–p)
- Displaced in the sense that the probability masses have moved k units to the right
- Conditioned on **X** = k, **Y** = k + **Z** = **X** + **Z** where **Z** is a Poisson random variable with parameter λ(1–p)

What's so important about all this?

- The example that we have studied arises in several different applications
- α-particles counted in a Geiger counter
- Each particle is detected with probability p and not detected with probability 1–p
- Detections are independent of each other
- If n α-particles are emitted, the number detected (the count in the Geiger counter) is a binomial RV with parameters (n,p)

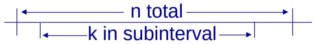
Demand better Geiger counters…

- α-particle emission is a Poisson process
- The number of α-particles emitted in unit time is a Poisson RV with parameter λ
- The number of α-particles detected in unit time is a Poisson RV with parameter λp
- If the Geiger counter counted k particles, what is the best estimate of how many particles were emitted?
- What is P{emissions = n | count = k} ??

Another application

- Consider a Poisson process with arrival rate μ
- Number of arrivals in interval of length T is a Poisson RV **Y** with parameter μ•T = λ
- The number of arrivals in a subinterval of length pT, 0 < p < 1, is a Poisson RV **X** with parameter μ•pT = λp
- Note that **Y** ≥ **X** always
- What is the joint pmf of **X** and **Y**?

Another application (continued)



n total

k in subinterval

- For n ≥ k, $p_{\mathbf{X},\mathbf{Y}}(k,n)$ = P{**X** = k, **Y** = n}

= P{k arrivals in subinterval, n arrivals total}

= P{k in subinterval, n–k in complement}

= P{k in subinterval}•P{n–k in complement}

since disjoint intervals are independent

= $\exp(-\lambda p)\bullet(\lambda p)^k/k!$

$\bullet\exp(-\lambda(1-p))\bullet(\lambda(1-p))^{n-k}/(n-k)!$

= $[n!/k!(n-k)!]\bullet p^k(1-p)^{n-k}\bullet\exp(-\lambda)\bullet\lambda^n/n!$

Example applies to Poisson process

- **X** = # arrivals in interval of length pT
- **Y** = # arrivals in longer interval of length T
- Joint pmf of **X** and **Y** is as in our example
- Conditional pmf of **X** given **Y** is binomial
- Conditional pmf of **Y** given **X** is displaced Poisson
- Given **X** = k, **Y**–k = **Z** = # arrivals in complementary interval is Poisson with parameter μ•(1–p)T = λ(1–p)

Summary

- We have studied joint pmfs of discrete random variables
- We have learned about the joint probability matrix and how to use it
- We have learned how to find marginal pmfs from joint pmfs
- We have learned about conditional pmfs
- We have learned a little more about Poisson processes