Discrete Random Variables

- Suppose that the discrete random variable \( X \) takes on values \( u_1, u_2, \ldots, u_m, \ldots \) 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 \( u_1, u_2, \ldots, u_m, \ldots \) and \( v_1, v_2, \ldots, v_m, \ldots \), respectively, is defined as:

\[
p_{X,Y}(u,v) = P\{X = u_i, Y = v_j\} \text{ if } u = u_i, v = v_j,
\]

and

\[
p_{X,Y}(u,v) = 0 \text{ otherwise.}
\]
- 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 is 1.
- Hence, \( \sum \sum p_{X,Y}(u_i, v_j) = 1 \).
- \( p_{X,Y}(u,v) \geq 0 \) for all \( u \) and \( v \), \( -\infty < u, v < \infty \).

Probabilities from the joint pmf:

- Let \( A \) denote a region of the plane.
- Then, \( P((X,Y) \in A) \) 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

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

A picture is worth a thousand words

- \( p_{X,Y}(u_i, v_j) \) is written as \( p(u_i, v_j) \) for brevity.
- The row and columns sum appear in the margins.
- Note the ordering of the \( v_j \)'s.

Marginal pmfs from the joint pmf

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

Why marginal, for crying out loud?

- 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 \( P(X = u_i, Y = v_j) \), i.e., the value of \( p_{X,Y}(u_i,v_j) \).

An example

<table>
<thead>
<tr>
<th>( u_1 )</th>
<th>( u_2 )</th>
<th>( u_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v_1 )</td>
<td>0.17</td>
<td>0.35</td>
</tr>
<tr>
<td>( v_2 )</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>( v_3 )</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>( v_4 )</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

- The marginal pmfs are as shown.
- \( P(X = Y) = 0.02 + 0.14 + 0.12 = 0.28 \).
- \( P(X + Y = 3) = 0.02 + 0.16 + 0 = 0.18 \).
Generalization: many discrete RVs

- Let $X_1, X_2, \ldots, X_n$ be $n$ discrete random variables defined on a sample space
- $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a random vector
- $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ is a real vector
- The notation $\{ \mathbf{X} = \mathbf{u} \}$ denotes the event $\{ X_1 = u_1, X_2 = u_2, \ldots, X_n = u_n \}$, where, as before, the commas denote intersections, that is, $\{ \mathbf{X} = \mathbf{u} \} = \{ X_1 = u_1 \} \cap \{ X_2 = u_2 \} \cap \ldots \cap \{ X_n = u_n \}$.

The joint pmf of many discrete RVs

- The joint pmf $p_{\mathbf{X}}(\mathbf{u}) = p(\mathbf{X} = \mathbf{u})$ of the random vector $\mathbf{X}$ is defined as $p_{\mathbf{X}}(\mathbf{u}) = p(X_1 = u_1, X_2 = u_2, \ldots, X_n = u_n)$.
- $p_{\mathbf{X}}(\mathbf{u}) \geq 0$
- $\sum_{u_1} \sum_{u_2} \ldots \sum_{u_n} p_{\mathbf{X}}(\mathbf{u}) = 1$
- The marginal pmf of any subset of $\{X_1, X_2, \ldots, X_n\}$ is obtained by summing over the unwanted variables.

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 $\{ \mathbf{X} = u_i \}$ from $P(\mathbf{B})$ to $P(\mathbf{B} | A)$.

Conditional pmfs, again?

- Given the event $\{ \mathbf{Y} = v_3 \} = A$ has occurred, we should update the probability of event $\mathbf{B} = (X = u_i)$ from $P(\mathbf{B})$ to $P(\mathbf{B} | A)$
- $P(\mathbf{B} | A) = P(\mathbf{X} = u_i, \mathbf{Y} = v_3) / P(\mathbf{Y} = v_3)$
- Remember, commas mean intersections
- $P(\mathbf{X} = u_i | \mathbf{Y} = v_3) = p_{X,Y}(u_i, v_3) / p_Y(v_3)$
- = 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 $p_{X,Y}(u_i | v_j) = p_{X,Y}(u_i, v_j) / p_Y(v_j)$
- Note that we are assuming that $p_Y(v_j) > 0$
- $p_{X,Y}(u_i | v_j) = p_{X,Y}(u_i | v_j)$ if $u_i$ is one of the values that $\mathbf{X}$ can take on
- $p_{X,Y}(u_i | v_j) = 0$ if $u_i$ is not any of the $u_i$

---

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

31.3
Conditional pmf is a function of u !!

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

Conditional pmf from the joint pmf

\[ p_{X|Y}(u|v_j) = \frac{p_{X,Y}(u, v_j)}{p_Y(v_j)} \]

- \( p_{X|Y}(u|v_j) \) has values 1/6, 1/3, 1/3 and 1/6 at u = –1, 0, 1, and 2
- \( p_{X|Y}(u|2) \) has values 0.01, 0.02, 0.02 and 0.01
- \( p_{X|Y}(u|3) \) has values 0.06, 0.44, 0.44 and 0.06

A numerical example

<table>
<thead>
<tr>
<th>u</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.17</td>
<td>0.35</td>
<td>0.34</td>
<td>0.14</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>1</td>
<td>0.15</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>0</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

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 \( \{Y = v_1\}, \{Y = v_2\}, \ldots, \{Y = v_m\}, \ldots \) are a partition of the sample space \( \Omega \)
- \( p_{X|Y}(u) = \sum_{j=1}^{m} p_{X,Y}(u, v_j) \)
- \( = \sum_{j=1}^{m} p_{X,Y}(u|v_j)p_Y(v_j) \)

Unconditional pmf from conditional

- Events \( \{Y = v_1\}, \{Y = v_2\}, \ldots, \{Y = v_m\}, \ldots \) are a partition of the sample space \( \Omega \)
- \( p_X(u) = \sum_{j=1}^{m} p_{X,Y}(u|v_j)p_Y(v_j) \)
- This is just the theorem of total probability
- \( P(B) = \sum P(B|A_i)P(A_i) \)
- expressed in terms of pmfs
Reversing the conditioning…

- \( p_X(u) = \sum p_{X,Y}(u,v) \)
- \( p_{Y|X}(v|u) = \frac{p_{X,Y}(u,v)}{p_X(u)} \)
- This is just Bayes’ formula

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, … \( n \)
- \( p_{X|Y}(k|n) = \binom{n}{k} p^k (1-p)^{n-k}, 0 \leq k \leq n \)

Example (continued)

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

The joint pmf of \( X \) and \( Y \)

- For \( 0 \leq k \leq n \), \( p_{X,Y}(k,n) = p_{X|Y}(k|n) \cdot p_Y(n) \)
- \( = \frac{n!}{k!} p^k (1-p)^{n-k} \cdot \exp(-\lambda) \cdot \frac{\lambda^n}{n!} \)
- The point masses lie in the triangular region shown

Some thoughts on the joint pmf

- \( p_{X,Y}(k,n) = \binom{n}{k} p^k (1-p)^{n-k} \cdot \exp(-\lambda) \cdot \frac{\lambda^n}{n!} \)
- Conditioned on \( Y = n \), \( X \) has values from 0 to \( n \). However, unconditionally, \( X \) takes on values 0, 1, 2, … the same as \( Y \) !!
- \( Y \geq X \) always

Unconditional (marginal) pmf of \( X \)

- \( p_X(k) = \sum_{n=k}^{\infty} p_{X,Y}(k,n) \)
- Where sum is over \( n \) from \( n = k \) to \( \infty \)
- \( p_X(k) = \exp(-\lambda p) \cdot (\lambda p)^k / k! \) for \( k = 0, 1, 2, \ldots \)
- \( X \) is a Poisson RV with parameter \( \lambda p \) !!
Some details of the calculations

- $p_X(k) = \sum p_{X|Y}(k,n)$
- $= \sum [p(k(n-k)!p^n(1-p)^{n-k}\exp(-\lambda)\lambda^n/n!]$
- Remember that the sum is over $n$ from $n = k$ to $\infty$
- $p_T(k) = \exp(-\lambda)\lambda^k/k!$
- The unconditional pmf of $X$ is Poisson with parameter $\lambda p$

What have we learned so far?

- $Y$ is a Poisson random variable with parameter $\lambda$
- 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
- $Y \geq X$ always
- The unconditional pmf of $X$ is Poisson with parameter $\lambda p$

Conditional pmf of $Y$ given $X$

- $p_{Y|X}(n|k) = p_{X|Y}(k,n)/p_X(k)$
- $= \exp(-\lambda(1-p))\lambda^{n-k}/(n-k)!$ for $n \geq k$
- $= \exp(-\lambda(1-p))\lambda^{n-k}/(n-k)!$ for $n \geq k$
- This is called a displaced Poisson pmf

Move to the right by $k$ places...

- For $n \geq k$, $p_{Y|X}(n|k) = p_{X|Y}(k,n)/p_X(k)$
- $= \exp(-\lambda(1-p))\lambda^{n-k}/(n-k)!$
- This is a displaced Poisson pmf with parameter $\lambda(1-p)$
- 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 $\lambda(1-p)$

What’s so important about all this?

- The example that we have studied arises in several different applications
- $\alpha$-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 $n \alpha$-particles are emitted, the number detected (the count in the Geiger counter) is a binomial RV with parameters $(n,p)$

Demand better Geiger counters...

- $\alpha$-particle emission is a Poisson process
- The number of $\alpha$-particles emitted in unit time is a Poisson RV with parameter $\lambda$
- The number of $\alpha$-particles detected in unit time is a Poisson RV with parameter $\lambda p$
- If the Geiger counter counted $k$ particles, what is the best estimate of how many particles were emitted?
- What is $P(\text{emissions} = n \mid \text{count} = k)$?
Another application

- Consider a Poisson process with arrival rate $\mu$.
- Number of arrivals in interval of length $T$ is a Poisson RV $Y$ with parameter $\mu T = \lambda$.
- The number of arrivals in a subinterval of length $pT$, $0 < p < 1$, is a Poisson RV $X$ with parameter $\mu pT = \lambda p$.
- Note that $Y \geq X$ always.
- What is the joint pmf of $X$ and $Y$?

Another application (continued)

For $n \geq k$, $p_{X,Y}(k,n) = P(X = k, Y = n) = P(k$ 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 = $\exp(-\lambda p)(\lambda p)^k/k!$ = $\exp(-\lambda(1-p))(\lambda(1-p))^{n-k}/(n-k)!$

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 $\mu(1-p)T = \lambda(1-p)$.

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.