May 27 Lec—Multinomial Cont., Conditional Expectation, Tower Property

Continuing from last time:

Proposition

If $X_1, X_2, \dots, X_k$ have a (joint) multinomial distribution with parameter $N, p_1, p_2, \dots, p_k$, then:

1. $V(X_i) = Np_i(1-p_i)$
2. $Cov(X_i, Xj) = -Np_ip_j$ if $i \neq j$

Proof

WLOG (without loss of generality), we'll prove that $Cov(X_1, X_2) = -Np_1p_2$.
For each experiment indexed with $k = 1, 2 \dots, N$, define

$$U_k = \begin{cases} 1 & \text{if the outcome is } w_1 \\ 0 & \text{otherwise} \end{cases}$$

$$V_k = \begin{cases} 1 & \text{if the outcome is } w_2 \\ 0 & \text{otherwise} \end{cases}$$

We quickly observe that $U_k \sim Ber(p_1), V_k \sim Ber(p_2)$.
Since X_1 is the total number of the occurrence of w_1, we have $X_1 = \sum_{k=1}^N U_k$ and $X_2 = \sum_{k=1}^N V_k$.
Since in a multinomial experiment the different trials are independent, we see that if $k \neq k'$, then U_k and $U_{k'}$, U_k and $V_{k'}$, and V_k and $V_{k'}$ are all independent in pairs.
So

$$Cov(x_1, x_2) = Cov\left(\sum_{k=1}^N U_k, \sum_{k'=1}^N V_{k'}\right)$$

$$= \sum_{k=1}^N \sum_{k'=1}^N Cov(U_k, V_{k'})$$

$$= \sum_{k=k'=1}^{N} Cov(U_k, V_{k'}) + \underbrace{\cancel{\sum_{k \neq k'} Cov(U_k, V_{k'})}}_{\text{by independence}}$$

$$= \sum_{k=1}^{N} Cov(U_k, V_k)$$

We calculate one term of $Cov(U_k, V_k)$:

$$Cov(U_k, V_k) = \mathbb{E}(U_k V_k) - \mathbb{E}(U_k)\mathbb{E}(V_k)$$
$$= \mathbb{E}(U_k V_k) - p_1 p_2$$

We notice that $U_k V_k = 0$ for all k (because it's either $1 \cdot 0$ or $0 \cdot 1$ or $0 \cdot 0$). So for all k,

$$Cov(U_k, V_k) = -p_1 p_2$$

It follows that

$$Cov(X_1, X_2) = \sum_{k=1}^{N} -p_1 p_2$$
$$= -N p_1 p_2$$

Example

An urn contains 3 Red balls, 5 Blue bals, and 2 Green balls.
Select $N = 10$ balls from the urn with replacement.
After each draw, the payoff is $+2$ for Red, -3 for Blue, $+1$ for Green.
Let Y be the total payoff. Find $\mathbb{E}(Y)$ and $V(Y)$.

Let X_R be the number of Reds, X_B be the number of Blues, and X_G be the number of Greens. X_R, X_B, X_G have a joint multinomial distribution with parameters $N = 10, p_R = \frac{3}{10}, p_B = \frac{5}{10}, p_G = \frac{2}{10}$
Then

$$Y = 2X_R - 3X_B + 1X_G$$
$$\implies \mathbb{E}(Y) = 2\mathbb{E}(X_R) - 3\mathbb{E}(X_B) + \mathbb{E}(X_G)$$
$$= 2 \cdot 10 \cdot \frac{3}{10} - 3 \cdot 10 \cdot \frac{5}{10} + 1 \cdot 10 \cdot \frac{2}{10} = -7$$

and

$$
\begin{aligned}
V(Y) &= Cov(Y,Y) \\
&= Cov(2X_R - 3X_B + 1X_G, 2X_R - 3X_B + 1X_G) \\
&= 4V(X_R) + 9V(X_B) + V(X_G) - 12Cov(X_R, X_B) + 4Cov(X_R, X_G) - 6Cov(X_B, X_G) \\
&= 4 \cdot 10 \cdot \frac{3}{10} \cdot \frac{7}{10} + 9 \cdot 10 \cdot \frac{5}{10} \cdot \frac{5}{10} + 10 \cdot \frac{2}{10} \cdot \frac{8}{10} - 12 \cdot (-10) \cdot \frac{3}{10} \cdot \frac{5}{10} + 4 \cdot (-10) \cdot \\
&= \frac{371}{10}
\end{aligned}
$$

(scribe's note: again, I am not a mathematician that knows how to count 😭)

Conditional Expectation

Motivating example:
Let X be the number of customers coming to a restaurant over a night, and suppose $X \sim P(\lambda)$. Each customer has a fixed probability of p of choosing Menu 1 and $1-p$ probability of choosing Menu 2. Let X_1 be the number of customers in total choosing Menu 1 and X_2 the number choosing Menu 2.
We'd be baffled at trying to find the distribution of X_1 and X_2. If we know the value of X (say, N) then $X_1 \sim B(N,p)$, but we don't know much beyond that. (It would be incorrect to say that $X_1 \sim B(X,p)$.)

Definition: Conditional expectations

1. Let (X,Y) be a pair of discrete random variables.
   Let $P(Y|X=x)$ be the conditional pf of Y given $X=x$.
   For any function $F: \mathbb{R} \to \mathbb{R}$, define the function $g(x) = \mathbb{E}(F(Y)|X=x)$, which is a function of X.
   The **conditional expectation** of $F(Y)$ given X is denoted $g(X) = \mathbb{E}(F(Y)|X)$ (a shorthand for $\mathbb{E}(F(Y)|X=x)$).
2. Similarly, if (X,Y) is a pair of continuous random variables, the definition of the conditional expectation is

$$g(x) = \mathbb{E}(F(Y)|X=x)$$

$$= \int_{\mathbb{R}} F(y) f(y|X=x)\, dy$$

Example

y**-value** x**-value**	-1	0	1	$p_Y(y)$
0	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{2}$
1	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
$p_X(x)$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	

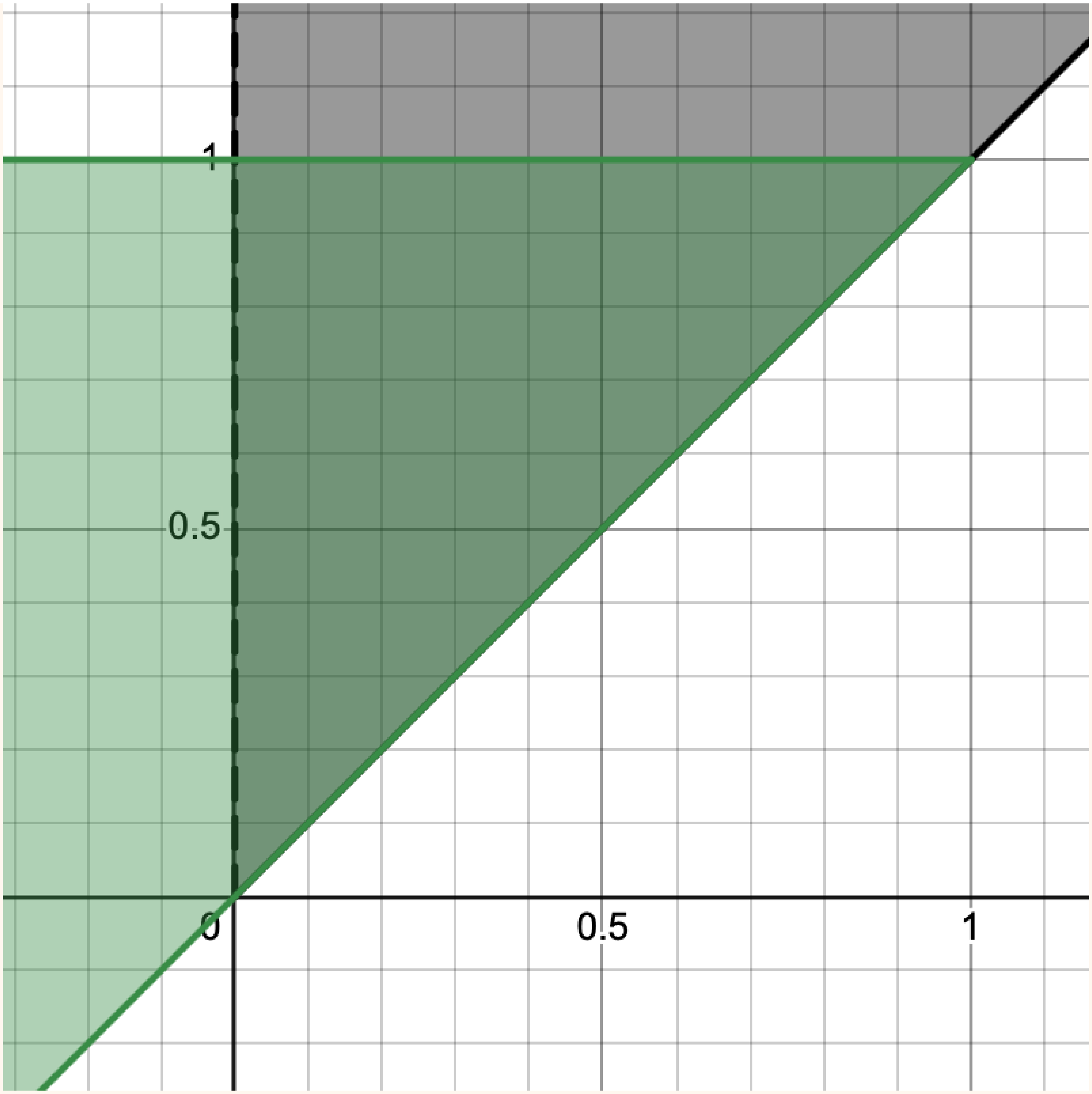
Find $g(X) := \mathbb{E}(Y|X)$ for all valid X.

$g(X)$ is defined only for $X = -1, 0, 1$.

- $g(-1) = \mathbb{E}(Y|X=-1)$. Given $X = -1$, we have that $Y \sim Ber\left(\frac{1}{2}\right)$, so $g(-1) = \frac{1}{2}$
- $g(0) = 0$
- $g(1) = \frac{3}{4}$

Continuous example

$$f(x,y) = \begin{cases} 6x & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$



We said many times that $X \sim Beta(2,2)$, $Y \sim Beta(3,1)$. Also, given $X = x$ we have $Y \sim U(x,1)$ and given $Y = y$ we found that $X \sim Y \cdot Beta(2,1)$.

Computing $\mathbb{E}(Y|X) =: g(X)$:
$g(X) = \mathbb{E}(Y|X=x) = \frac{1+x}{2}$ since $Y \sim U(x,1)$
We can also compute $h(X) := \mathbb{E}(Y^2|X)$:
$\mathbb{E}(Y^2|X) = \frac{(1-x)^2}{12} + \left(\frac{1+x}{2}\right)^2$ (this is what we did in the proof of uniform distribution variance)
Similarly, $g_1(Y) := \mathbb{E}(X|Y) = y \cdot \frac{2}{3}$
For $h_1(Y) := \mathbb{E}(X^2|Y)$
For fixed Y, $X = Y \cdot W$ where $W \sim Beta(2,1) \implies X^2 = Y^2 \cdot W^2$.

$$\implies \mathbb{E}(X^2|Y)$$

A third example

Continuing the restaurant example we had above.
What is $\mathbb{E}(X_1|X)$ and $\mathbb{E}(X_2|X)$?
We already said that $\mathbb{E}(X_1|X=N)=Np$ (since we said before that given $X=N$ we have that $X_1 \sim B(N,p)$) and similarly $\mathbb{E}(X_2|X=N)=N(1-p)$.

Properties of conditional expectations

1. $\mathbb{E}(F(y)+G(y)|X)=\mathbb{E}(F(y)|X)+\mathbb{E}(G(y)|X)$
2. If F and G are two functions, then $\mathbb{E}(F(X)G(Y)|X)=F(X)\mathbb{E}(G(Y)|X)$.
3. If X and Y are independent, $\mathbb{E}(F(Y)|X)=\mathbb{E}(F(Y))$.
4. **Tower Property:** $\mathbb{E}(\mathbb{E}(F(Y)|X))=\mathbb{E}(F(Y))$

Proof

1 and 2 should be straightforward enough.

4. Let $G(x):=\mathbb{E}(F(Y)|X)$; we want to compute

$$\int_{-\infty}^{\infty} G(X) f_X(x)\,dx$$

We have

$$G(X)=\int_{-\infty}^{\infty} F(y) f(y|X=x)\,dy$$

so

$$\mathbb{E}(\mathbb{E}(F(Y)|X))=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} F(y) f(y|X=x)\,dy\right) f_X(x)\,dx$$

Since the bounds of integration are both $(-\infty,\infty)$, we can switch the order of integration without any qualms.

$$\mathbb{E}(\mathbb{E}(F(Y)|X))=\int_{-\infty}^{\infty} F(y) \underbrace{\left(\int_{-\infty}^{\infty} \underbrace{f(y|X=x) f_X(x)}_{f(x,y)}\,dx\right)}_{f_Y(y)} dy$$

$$= \int_{-\infty}^{\infty} F(y) f_Y(y)\, dy$$

$$= \mathbb{E}(F(y))$$

5. Assume X, Y are independent.

$$\mathbb{E}(F(Y)|X = x) = \int_{-\infty}^{\infty} F(y) \underbrace{f(y|X = x)}_{= f_Y(y) \text{ by independence}} dy$$

$$= \mathbb{E}(F(y))$$

We can view $\mathbb{E}(Y|X)$ as the orthogonal projection of Y on the space of the functions of X.

Proposition

$Y - \mathbb{E}(Y|X)$ is uncorrelated with every function of X.

Proof

We want to show that $Cov(Y - \mathbb{E}(Y|X), H(X)) = 0$ for every $H(x)$.
We know by the Tower Property $\mathbb{E}(Y - \mathbb{E}(Y|X)) = 0$
We have

$$Cov(Y - \mathbb{E}(Y|X), H(X)) = \mathbb{E}(YH(X) - \mathbb{E}(Y|X)H(X)$$

$$Cov(Y - \mathbb{E}(Y|X), H(X)) = \mathbb{E}((Y - \mathbb{E}(Y|X)) \cdot H(X)) - \cancel{\mathbb{E}(Y - \mathbb{E}(Y|X)) \cdot \mathbb{E}(H(X))}$$

Example

$$f(x, y) = \begin{cases} 6x & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

and $X \sim Beta(2, 2)$, $Y \sim Beta(3, 1)$
Given $X = x$, $Y \sim U(x, 1)$
Given $Y = y$, $X = yW$ where $W \sim Beta(2, 1)$
We computed above that $\mathbb{E}(Y|X) = \frac{1+x}{2}$. Then,

$\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}\left(\frac{1}{2} + \frac{x}{2}\right) = \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4} = \mathbb{E}(Y)$ as we expected.
Also, $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}\left(y \cdot \frac{2}{3}\right) = \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{2} = \mathbb{E}(X)$ as expected.

Continuing the restaurant example above...

What is the mgf of X_1?

$$m_{X_1}(t) = \mathbb{E}(e^{tX_1})$$

By the Tower Property,

$$= \mathbb{E}(\mathbb{E}(e^{tX_1}|X))$$

We already know that given $X = N$, $X_1 \sim B(N, p)$. Therefore, by the formulas we had before,

$$H(X) := \mathbb{E}(e^{tX_1}|X) = [pe^t + (1-p)]^X$$

and hence

$$m_{X_1}(t) = \mathbb{E}(e^{tX_1})$$

$$= \mathbb{E}(H(X))$$

$$= \mathbb{E}([pe^t + (1-p)]^X)$$

If $X \sim P(\lambda)$,
and defining $a := pe^t + (1-p)$

$$m_{X_1}(t) = \mathbb{E}(e^{X \ln a})$$

$$= m_X(t = \ln(a))$$

$$e^{\lambda(e^{\ln a} - 1)}$$

$$= e^{\lambda(a-1)}$$

So

$$m_{X_1}(t) = e^{\lambda(pe^t + (1-p) - 1)}$$

$$= m_{X_1}(t) = e^{\lambda p(e^t - 1)}$$

So (drum rolls!!!)

$$X_1 \sim P(\lambda p)$$

Aside: a special challenge

Prove that X_2 and X_1 are independent in the above example.