Stats 426 Lecture 5

Thomas Cohn

2/3/21

Defn: The multinomial distribution is a multivariate version of the binomial distribution, where there are n independent trials, and each trial results in 1 of r different outcomes, with probabilities $p_1, \dots, p_r$ s.t. $p_1 + \cdots + p_r = 1$. If $(X_1, \dots, X_r)$ is a vector of binomial random variables, then the pmf of the multinomial distribution is

$$p(n_1, \dots, n_r) = \binom{n}{n_1, \dots, n_r} p_1^{n_1} \cdots p_r^{n_r}$$

when the n_i's sum to n, and 0 otherwise. Note that the pmf is defined in terms of the multinomial coefficient

$$\binom{n}{n_1, \dots, n_r} \overset{\text{def}}{=} \frac{n!}{n_1! \cdots n_r!}$$

Defn: Let X and Y be continuous random variables. Then $f(x,y)$ is the joint probability density function of (X,Y) if, for any $A \subseteq \mathbb{R}^2$,

$$P((x,y) \in A) = \iint_A f(x,y)\,dx\,dy$$

This joint pdf must satisfy

- $f(x,y) \geq 0$, $\forall (x,y)$
- $\displaystyle\iint_{\mathbb{R}^2} f(x,y)\,dx\,dy$

Defn: The marginal probability distrbutions of X and Y are

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy \qquad f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx$$

(respectively).

Defn: X and Y are independent if, $\forall (x,y)$, $f(x,y) = f_X(x) f_Y(y)$.

Defn: The joint cumulative distribution function is

$$F(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\,dv\,du$$

Ex: Let (X,Y) have the joint pdf

$$f(x,y) = \begin{cases} \frac{4}{5}xy & 0 \leq x \leq 1, 2 \leq y \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

Then the marginal pdfs are

$$f_X(x) = \int_2^3 \frac{4}{5}xy\,dy = \left[\frac{2}{5}xy^2\right]_{y=2}^{y=3} = \frac{2}{5}x(9-4) = 2x \qquad \text{(for } 0 \leq x \leq 1\text{)}$$

$$f_Y(y) = \int_0^1 \frac{4}{5}xy\,dx = \left[\frac{2}{5}x^2 y\right]_{x=0}^{x=1} = \frac{2}{5}y(1-0) = \frac{2}{5}y \qquad \text{(for } 2 \leq y \leq 3\text{)}$$

Are X and Y independent? Yes! $f_X(x)f_Y(y) = \frac{4}{5}xy = f(x,y)$, $\forall (x,y)$. We write "$X \perp Y$".

Ex: A point is chosen randomly in a disc of radius 1. Let X, Y be its x, y coordinates. The pdf of X, Y is

$$f(x,y) = \begin{cases} \frac{1}{\pi} & x^2+y^2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

What are f_X and f_Y?

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\, dy = \frac{2}{\pi}\sqrt{1-x^2} \qquad x \in [-1,1]$$

$$f_Y(y) = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{1}{\pi}\, dx = \frac{2}{\pi}\sqrt{1-y^2} \qquad y \in [-1,1]$$

$f_X(x)f_Y(y) = \frac{4}{\pi^2}\sqrt{1-x^2}\sqrt{1-y^2} \neq f(x,y)$ in general, so X and Y are not independent.

The Bivariate Normal Distribution

Defn: (X,Y) follows a bivariate normal distribution with parameters $\mu_X, \mu_Y, \sigma_X, \sigma_Y > 0$, and $\rho \in (-1,1)$ if the pdf is

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)\right]}$$

The special case where $\rho = 0$ is the case where $X \perp Y$.

Recall: For 2 independent random variables X and Y, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$.
Random variables $X_1, \dots, X_n$ are independent when $f(x_1,\dots,x_n) = f(x_1)\cdots f(x_n)$, $\forall (x_1,\dots,x_n)$, which is true when $F(x_1,\dots,x_n) = F(x_1)\cdots F(x_n)$, $\forall (x_1,\dots,x_n)$.

Conditional Distributions

Defn: For discrete random variables X and Y, the conditional probability density function of X given Y is

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

when $p_Y(y) > 0$, and 0 otherwise. That is, $p_{X,Y}(x,y) = p_{X|Y}(x \mid y)p_Y(y)$.

Consequently, the marginal distribution of X is $p_X(x) = \sum_y p_{X|Y}(x \mid y)p_Y(y)$.

Defn: For continuous random variables X and Y, the conditional probability density function of X given Y is

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

when $f_Y(y) > 0$, and 0 otherwise.

Ex: Say (X,Y) follow a bivariant normal distribution with parameters $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$. The conditional density of X given $Y = y$ is

$$\mathrm{N}\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y), (1-\rho^2)\sigma_X^2\right)$$

Observe that the variance of this new distribution is no more than that of X.

Ex: Say $X_1 \sim \text{Unif}[0,1]$, and $X_2 \sim \text{Unif}[0, X_1]$. We want to find the joint pdf of (X_1, X_2), and the marginal pdf of X_2. Well, $f_{X_1} = 1$ and $f_{X_2|X_1} = \frac{1}{X_1}$. So $f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2) = \frac{1}{x_1}$, for $0 \le x_1 \le 1$, and $0 \le x_2 \le x_1$ (and 0 otherwise). Thus,

$$f_{X_2}(x_2) = \int_{x_2}^{1} \frac{1}{x_1}\, dx_1 = [\ln(x_1)]_{x_1=x_2}^{x_1=1} = -\ln x_2 \qquad 0 < x_2 \le 1$$

Ex: (Bayesian Posterior)

Given $0 < \theta < 1$, suppose $X \sim \text{Bin}(n, \theta)$. Suppose the prior pdf of Θ is Unif$[0,1]$. What is $p_X(x)$?

Well, $f(x,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ when $x = 0, 1, \ldots, n$, and $0 < \theta < 1$. So we can compute the marginal pdf of X:

$$\begin{aligned}
p_X(x) &= \int_0^1 \binom{n}{x}\theta^x(1-\theta)^{n-x}\,d\theta \\
&= \binom{n}{x}\int_0^1 \theta^x(1-\theta)^{n-x}\,d\theta \\
&= \binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+x)}\underbrace{\int_0^1 \frac{\Gamma(x+1+n-x+1)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^{(x-1)+1}(1-\theta)^{(n-x+1)-1}\,d\theta}_{\text{The pdf of Beta}(x+1,\,n-x+1)} \\
&= \binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+x)}
\end{aligned}$$

Now, show that the posterior pdf of Θ given $X = x$ is

$$\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x}$$

which is a beta distribution. Well

$$f(x,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}; \quad f_X(x) = p_X(x) = \binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}$$

so

$$f_{\Theta|X}(\theta) = \frac{f(x,\theta)}{f_X(x)} = \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x}}{\binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}} = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x} \sim \text{Beta}(x+1, n-x+1)$$

$\square$