14.381 HANDOUT #3

10/15/2004

FALL 2004

TA: JOSÉ TESSADA

1. MARGINAL AND CONDITIONAL DISTRIBUTIONS

Definition: Let (X, Y) be a bivariate random vector with joint pmf (pdf) $f_{X,Y}(x, y)$. Then the marginal pmfs (pdfs) of X and Y, $f_X(x)$ and $f_Y(y)$, are given by

$$f_X(x) = \begin{cases} \sum_{y \in \mathbb{R}} f_{X,Y}(x, y) & \text{if discrete} \\ \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy & \text{if continuous} \end{cases}$$

$$f_Y(y) = \begin{cases} \sum_{x \in \mathbb{R}} f_{X,Y}(x, y) & \text{if discrete} \\ \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx & \text{if continuous} \end{cases}$$

Definition: Let (X, Y) be a discrete bivariate random vector with joint pmf $f(x, y)$ and marginal pmfs $f_X(x)$ and $f_Y(y)$. For any x such that $P(X=x) = f_X(x) > 0$, the conditional pmf of Y given $X=x$ is the function of y denoted by $f(y|x)$ and defined by

$$f(y|x) = P(Y=y \mid X=x) = \frac{f(x, y)}{f_X(x)}$$

Notes:
- definition is analogous for $f(x|y)$;
- conditional pmfs are well defined pmfs:
  - $f(y|x) \geq 0 \quad \forall y$ since $f(x,y) \geq 0$ and $f_X(x) > 0$
  - $$\sum_{y \in \mathbb{R}} f(y|x) = \left[\sum_{y \in \mathbb{R}} f(x, y)\right] \cdot \frac{1}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1$$

<u>Definition</u>: Let (X,Y) be a continuous bivariate random vector with joint pdf $f(x,y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$. For any x such that $f_X(x)>0$, the conditional pdf of Y given that $X=x$ is the function of y denoted by $f(y|x)$ and defined by

$$f(y|x) = \frac{f(x,y)}{f_X(x)}$$

Remember that conditional distributions must be interpreted with caution in the continuous case! (see section 4.9.3)

<u>Exercise 4.10</u>: the random pair (X,Y) has the distribution

		$X=1$	$X=2$	$X=3$
Y	2	1/12	1/6	1/12
	3	1/6	0	1/6
	4	0	1/3	0

(a) Show that X and Y are dependent:

The marginal distribution of X is

$$P(X=1) = P(X=3) = 1/4 \quad \text{and} \quad P(X=2) = 1/2.$$

$$P(X=x) = \sum_{y=2,3,4} P(x,y) \qquad \text{ex.: } P(X=1) = P(X=1, Y=2) + P(X=1, Y=3) + P(X=1, Y=4) = 1/12 + 1/6 + 0$$

The marginal distribution of Y is

$$P(Y=2) = P(Y=3) = P(Y=4) = 1/3$$

It suffices to show that

$$P(X=2, Y=3) = 0 \neq \frac{1}{2}\cdot\frac{1}{3} = P(X=2)\cdot P(Y=3)$$

so X and Y are dependent.

(b) Give a probability table for random variables U and V that have the same marginals as X and Y but are independent.

We need a bivariate distribution such that $P(U=x, V=y) = P(U=x)\,P(V=y)$, where $U \sim X$ and $V \sim Y$. The following distribution does that:

	X: 1	2	3
V: 2	1/12	1/6	1/12
3	1/12	1/6	1/12
4	1/12	1/6	1/12

Exercise 4.15: Let $X \sim \text{Poisson}(\theta)$, $Y \sim \text{Poisson}(\lambda)$, independent. From theorem 4.3.2: $X+Y \sim \text{Poisson}(\theta+\lambda)$. Show that the distribution of $Y \mid X+Y$ is binomial with success probability $\lambda/(\theta+\lambda)$.

Let $U = X+Y$ and $V = Y$. the joint pmf of (U,V) is

$$f_{U,V}(u,v) = f_{X,Y}(u-v, v) = \frac{\theta^{u-v} e^{-\theta}}{(u-v)!} \frac{\lambda^{v} e^{-\lambda}}{v!}$$

This is true because for any v and $u\ (= x+y)$, only $x = u-v$ is possible.

Now notice that for fixed u, $f(u,v)$ is positive for $v = 0, \dots, u$. Therefore the conditional pmf is

$$f(v \mid u) = \frac{f(u,v)}{f(u)} = \frac{\dfrac{\theta^{u-v} e^{-\theta}}{(u-v)!} \dfrac{\lambda^{v} e^{-\lambda}}{v!}}{\dfrac{(\theta+\lambda)^{u}\, e^{-(\theta+\lambda)}}{u!}}$$

$$= \frac{u!}{(u-v)!\,v!} \cdot \frac{e^{-\theta}\, e^{-\lambda}}{e^{-(\theta+\lambda)}} \cdot \frac{\theta^{u-v} \lambda^{v}}{(\theta+\lambda)^{u}}$$

$$f(v \mid u) = \binom{u}{v} \left(\frac{\theta}{\theta+\lambda}\right)^{u-v} \left(\frac{\lambda}{\theta+\lambda}\right)^{v} \qquad v = 0, 1, \dots, u$$

2. INDEPENDENCE

Def. 4.2.5 Let (X,Y) be a bivariate random vector with joint pdf or pmf $f(x,y)$ and marginal pdfs or pmfs $f_X(x)$ and $f_Y(y)$. Then X and Y are called independent random variables if, for every $x \in \mathbb{R}$ and $y \in \mathbb{R}$,
$$f(x,y) = f_X(x) f_Y(y).$$

$$\Rightarrow f(y|x) = f_Y(y)$$

Lemma 4.2.7 Let (X,Y) be a bivariate random vector with joint pdf or pmf $f(x,y)$. Then X and Y are independent random variables if and only if there exist functions $g(x)$ and $h(y)$ such that, for every $x \in \mathbb{R}$ and $y \in \mathbb{R}$
$$f(x,y) = g(x) h(y).$$

Example 4.2.8
$$f(x,y) = \frac{1}{384} x^2 y^4 e^{-y-(x/2)}, \quad x>0, \ y>0$$

we can write
$$f(x,y) = x^2 e^{-x/2} \mathbb{1}(x>0) \ \frac{y^4 e^{-y}}{384} \cdot \mathbb{1}(y>0)$$
$$= g(x) \cdot h(y)$$
$\Rightarrow$ X and Y are independent.

From handout #2:

$$f(x,y) = \frac{1}{x} \qquad 0 \le y \le x \le 1$$

we can write $f(x,y) = \frac{1}{x} \cdot \mathbb{1}(y \in [0,x])$

$\Rightarrow$ we cannot write $f(\cdot) = g(x) \cdot h(y)$

$\Rightarrow$ X and Y are not independent.

<u>Definition 4.6.5</u> Let $X_1, \dots, X_n$ be random vectors with joint pdf or pmf $f(X_1, \dots, X_n)$. Let $f_{X_i}(x_i)$ denote the marginal pdf or pmf of X_i. Then $X_1, \dots, X_n$ are called mutually independent random vectors if, $\forall (x_1, \dots, x_n)$,

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

when X_i's are random variables, we call them mutually independent random variables.

<u>Theorem 4.6.11</u> Let $X_1, \dots, X_n$ be random vectors. Then $X_1, \dots, X_n$ are mutually independent random vectors iff $\exists$ functions $g_i(x_i)$, $i = 1, \dots, n$, such that the joint pdf or pmf of $(X_1, \dots, X_n)$ can be written as

$$f(x_1, \dots, x_n) = g_1(x_1) \cdot \dots \cdot g_n(x_n)$$

Notes:
- mutual independence implies more than pairwise independence, but pairwise independence does not imply mutual independence;
- mutual independence implies that knowing the values of some of the coordinates conveys no information about the values of other coordinates (conditional distribution is the same as the marginal).

3. CONDITIONAL MOMENTS

Thm. 4.4.3 Law of Iterated Expectations
If X and Y are any two random variables, then
$EX = E(E(X|Y))$,
provided that the expectation exist.

Example 4.4.1 - 4.4.2

$X|Y \sim \text{binomial}(Y, p)$
$Y \sim \text{Poisson}(\lambda)$

$$P(X=x) = \sum_{y=0}^{\infty} P(X=x, Y=y) = \sum_{y=0}^{\infty} P(X=x \mid Y=y) \cdot P(Y=y)$$

$$= \sum_{y=x}^{\infty} \underbrace{\left[\binom{y}{x} p^x (1-p)^{y-x}\right]}_{\text{binomial}(y,p)} \underbrace{\left[\frac{e^{-\lambda}\lambda^y}{y!}\right]}_{\text{Poisson}(\lambda)}$$

Note: conditional probability is 0 if $y < x$.

$$P(X=x) = \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{((1-p)\lambda)^{y-x}}{(y-x)!}$$

$$= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{t=0}^{\infty} \frac{((1-p)\lambda)^t}{t!} \qquad (\text{define } t = y-x)$$

$$= \frac{(\lambda p)^x e^{-\lambda}}{x!} e^{(1-p)\lambda} \qquad \left(\text{you can use the Poisson pmf or the def. of } e^a\right)$$

$$= \frac{(\lambda p)^x e^{-\lambda p}}{x!}$$

$\Rightarrow X \sim \text{Poisson}(\lambda p)$.

Now we know that $E(X) = \lambda p$ by the properties of the Poisson distribution.

However we could have used Thm. 4.4.3 in the following way:

$$E(X|Y) = pY \qquad \text{(property of binomial distribution}$$

and by iterated expectations

$$E(X) = E(E(X|Y))$$
$$= E(pY) = p \cdot E(Y) = p \cdot \lambda$$

$$\Rightarrow E(X) = p \cdot \lambda$$

It is useful to think about the distributions we use when computing the expectations:

$$E(X|Y) = \int_{-\infty}^{\infty} x \, f_{X|Y}(x|y) \, dx$$

so here we use the conditional distribution of X given $Y=y$.

Now, keep in mind that $E(X|Y)$ is not a function of X but of Y $\Rightarrow$

$$\int_{-\infty}^{\infty} E(X|Y) f_X(x) \, dx = E(X|Y)$$

So, we need to integrate over Y.
As a shortcut, let $E_Y(g(Y)) = \int_{-\infty}^{\infty} g(y) f_Y(y) \, dy$

$$\Rightarrow \quad E_X(X) = E_Y(E_{X|Y}(X|Y)).$$

<u>Thm. 4.4.7</u> Conditional Variance Identity
For any two random variables X and Y,
$Var(X) = E(Var(X|Y)) + Var(E(X|Y))$,
provided that the expectations exist.

(For an application see Example 4.4.8 in C&B).

4. COVARIANCE / CORRELATION

Def. 4.5.1-2

(1) The covariance of X and Y is the number defined by

$$Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y)).$$

(2) The correlation of X and Y is the number defined by

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y},$$

a.k.a. correlation coefficient

Thm. 4.5.5: If X and Y are independent random variables, then $Cov(X,Y)=0$ and $\rho_{XY}=0$.

Note: Jumping ahead ... $X \perp\!\!\!\perp Y \Rightarrow Cov(X,Y)=0$ but $Cov(X,Y) \not\Rightarrow X \perp\!\!\!\perp Y$ unless X and Y follow a bivariate normal distribution.

Example 4.5.9 $\longrightarrow$ shows a case where a pair of variables have a strong connection, but the nonlinearity leads to $Cov(\,)=0$.

$X \sim U(-1,1) \qquad Z \sim U(0, 1/10)$

$Y = X^2 + Z$

$$Y|X \sim U\left(x^2, x^2+\tfrac{1}{10}\right)$$

$$f(x,y) = 5, \qquad -1<x<1, \quad x^2<y<x^2+1/10$$

and $Cov(X,Y) = \rho_{XY} = 0$! And it should be clear that the variables are not independent (by simple inspection we know the joint pdf cannot be factorized).

5. BIVARIATE NORMAL DISTRIBUTION

Definition: Let $-\infty < \mu_x < \infty$, $-\infty < \mu_y < \infty$, $0 < \sigma_x, \sigma_y$, and $\rho \in (-1, 1)$ be real numbers. The bivariate normal pdf with means μ_x and μ_y, variances σ_x^2 and σ_y^2, and correlation ρ is given by

$$f(x,y) = \frac{1}{2\pi} \cdot \frac{1}{|\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (z-\mu)' \Sigma^{-1} (z-\mu) \right\}$$

where $z = (x \;\; y)'$, $\mu = (\mu_x \;\; \mu_y)'$ and

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

Some results from linear algebra:

① $\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = |A_{22}|\,|A_{11} - A_{12}A_{22}^{-1}A_{21}| = |A_{11}|\,|A_{22} - A_{21}A_{11}^{-1}A_{12}|$

② $\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} F_1 & -F_1 A_{12} A_{22}^{-1} \\ -F_2 A_{21} A_{11}^{-1} & F_2 \end{pmatrix} = \begin{pmatrix} F_1 & -A_{11}^{-1} A_{12} F_2 \\ -A_{22}^{-1} A_{21} F_1 & F_2 \end{pmatrix}$

where $F_1 = \left[A_{11} - A_{12} A_{22}^{-1} A_{21} \right]^{-1}$

$F_2 = \left[A_{22} - A_{21} A_{11}^{-1} A_{12} \right]^{-1}$

③ $\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{bmatrix} F_1 & -F_1 A_{12} A_{22}^{-1} \\ -(A_{12}A_{22}^{-1})' F_1 & A_{22}^{-1} + (A_{12}A_{22}^{-1})' F_1 (A_{12}A_{22}^{-1}) \end{bmatrix}$

Theorem: if (X,Y) has bivariate normal distribution, then the marginal distribution of X is $N(\mu_x, \sigma_x^2)$ and that of Y is $N(\mu_y, \sigma_y^2)$
[Note: the converse is not true.]

Theorem: for a bivariate normal, the conditional density of Y given $X=x$ is univariate normal with mean $\mu_y + (\rho\sigma_y/\sigma_x)(x-\mu_x)$ and variance $\sigma_y^2(1-\rho^2)$. The conditional density of X given $Y=y$ is also normal with mean $\mu_x + (\rho\sigma_x/\sigma_y)(y-\mu_y)$ and variance $\sigma_x^2(1-\rho^2)$.

For the bivariate case we can write:

$$f_{X,Y}(x,y) = \frac{1}{2\pi}\,\frac{1}{\sigma_x\sigma_y(1-\rho^2)^{1/2}}\exp\left\{-\frac{1}{2}\left(\frac{1}{1-\rho^2}\right)\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}$$

We want to rewrite it in this way: $f(x|y)\cdot f(y)$

where $f(y) = \frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sigma_2}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}$

$$\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2$$

$$= \left(\frac{x-\mu_x}{\sigma_x} - \rho\left(\frac{y-\mu_y}{\sigma_y}\right)\right)^2 + (1-\rho^2)\left(\frac{y-\mu_y}{\sigma_y}\right)^2$$

plugging into $f_{X,Y}(x,y)$:

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}}\,\frac{1}{\sigma_y}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}\cdot\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sigma_x(1-\rho^2)^{1/2}}\exp\left\{\frac{-1}{2}\,\frac{1}{1-\rho^2}\left[\frac{x-\mu_x}{\sigma_x} - \rho\left(\frac{y-\mu_y}{\sigma_y}\right)\right]^2\right\}$$

so,

$$f_{X,Y}(x,y) = g(y) \cdot h(x,y)$$

where $g(y)$ is the pdf of $N(\mu_Y, \sigma_Y^2)$ and $h(x,y)$ is the pdf of $N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}\cdot(y-\mu_Y),\ \sigma_X^2(1-\rho^2)\right)$

so $X \mid Y=y \sim N\left(\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y-\mu_Y),\ \sigma_X^2(1-\rho^2)\right)$

Now a more formal proof of the theorem for the marginals requires the following:

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y(1-\rho^2)^{1/2}} \exp\left\{-\frac{1}{2}\frac{1}{(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\} dy$$

Substitute $z = \frac{y-\mu_Y}{\sigma_Y}$, $dy = \sigma_Y\, dz$, $w = \frac{x-\mu_X}{\sigma_X}$

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y(1-\rho^2)^{1/2}} \exp\left\{-\frac{1}{2}\frac{1}{(1-\rho^2)}(w^2 - 2\rho wz + z^2)\right\}\sigma_Y\, dz$$

$$f_X(x) = \frac{\exp\left\{\frac{-w^2}{2(1-\rho^2)}\right\}}{2\pi\sigma_X(1-\rho^2)^{1/2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\frac{1}{(1-\rho^2)}\left[(z^2 - 2\rho wz + \rho^2 w^2) - \rho^2 w^2\right]\right\} dz$$

$$f_X(x) = \frac{\exp\left\{-w^2/2(1-\rho^2) + \rho^2 w^2/2(1-\rho^2)\right\}}{2\pi\sigma_X(1-\rho^2)^{1/2}} \underbrace{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\frac{1}{1-\rho^2}(z-\rho w)^2\right\} dz}_{A}$$

we can complete the pdf of a $N(\rho w, (1-\rho^2))$ in A, thus

$$f_X(x) = \frac{\exp\{-w^2/2\}}{2\pi\sigma_X\sqrt{1-\rho^2}} \sqrt{2\pi}\sqrt{1-\rho^2} = \frac{1}{\sqrt{2\pi}\,\sigma_X}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\}$$

which is the pdf of $N(\mu_X, \sigma_X^2)$. ∎

Note: • Let $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, that doesn't imply $(X, Y)'$ is distributed bivariate normal.

• for the bivariate normal, $Cov(X,Y) = 0$ (which is equivalent to say $\rho_{XY} = 0$) implies $X \perp\!\!\!\perp Y$. You can see why this is true by the fact that you cannot write $f(x,y) = g(x)h(y)$ unless $\rho = 0$.