Math 280B, Winter 2005

Conditioning and the Bivariate Normal Distribution

In what follows, X and Y are random variables defined on a probability space $(\Omega, \mathcal{B}, P)$, and $\mathcal{G}$ is a sub-σ-field of $\mathcal{B}$.

1. Regular Conditional Distributions. The conditional probability $P[X \in B|\mathcal{G}]$ is defined to be the conditional expectation $E[1_{\{X\in B\}}|\mathcal{G}] = E[1_B(X)|\mathcal{G}]$, for $B \in \mathcal{B}_{\mathbf{R}}$. The function $B \mapsto P[X \in B|\mathcal{G}]$ is "almost" a probability measure, in that $P[X \in \mathbf{R}|\mathcal{G}] = 1$ almost surely and $P[X \in \cup_{n=1}^\infty B_n\,|\mathcal{G}] = \sum_{n=1}^\infty P[X \in B_n|\mathcal{G}]$ almost surely for each sequence $\{B_n\}$ of pairwise disjoint elements of $\mathcal{B}_{\mathbf{R}}$. The ambiguity present in these "almost surely" statements can be resolved because X is real-valued. (Similar considerations apply to a random variable with values in a measurable space that is measurably isomorphic to $(\mathbf{R}, \mathcal{B}_{\mathbf{R}})$.) This resolution permits a converse linkage between conditional probabilities and conditional expectations. The situation is summarized in the following result.

Theorem. *There is a function $(\omega, B) \mapsto Q(\omega, B)$ from $\Omega \times \mathcal{B}_{\mathbf{R}}$ to $[0,1]$ such that (i) $\omega \mapsto Q(\omega, B)$ is $\mathcal{G}$-measurable for each $B \in \mathcal{B}_{\mathbf{R}}$, (ii) $B \mapsto Q(\omega, B)$ is a probability measure on $(\mathbf{R}, \mathcal{B}_{\mathbf{R}})$ for each $\omega \in \Omega$, and (iii) for each $B \in \mathcal{B}_{\mathbf{R}}$,*

$$P[X \in B|\mathcal{G}](\omega) = Q(\omega, B) \quad \text{for } P\text{-a.e. } \omega \in \Omega. \tag{1.1}$$

Moreover, if $\varphi : \mathbf{R} \to \mathbf{R}$ is Borel measurable and $E|\varphi(X)| < \infty$, then

$$E[\varphi(X)|\mathcal{G}](\omega) = \int_{\mathbf{R}} \varphi(x)\, Q(\omega, dx) \quad \text{for } P\text{-a.e. } \omega \in \Omega. \tag{1.2}$$

The function $Q(\omega, B)$ is called a regular conditional distribution for X given $\mathcal{G}$.

When $\mathcal{G}$ is of the form $\sigma(Y)$ there is a function $(y, B) \mapsto F_y(B)$ from $\mathbf{R} \times \mathcal{B}_{\mathbf{R}}$ to $[0,1]$ such that (i) $y \mapsto F_y(B)$ is $\mathcal{B}_{\mathbf{R}}$-measurable for each $B \in \mathcal{B}_{\mathbf{R}}$, (ii) $B \mapsto F_y(B)$ is a probability measure on $(\mathbf{R}, \mathcal{B}_{\mathbf{R}})$ for each $\omega \in \Omega$, and (iii) $Q(\omega, B) := F_{Y(\omega)}(B)$ is a regular conditional distribution for X given $\sigma(Y)$. It is natural to interpret $F_y(B)$ as the conditional probability $P[X \in B|Y = y]$. A parallel interpretation of (1.2) is

$$E[\varphi(X)|Y = y] = \int_{\mathbf{R}} \varphi(x)\, F_y(dx). \tag{1.3}$$

2. Basic Definition. A pair (X, Y) of random variables, defined on some probability space $(\Omega, \mathcal{F}, P)$, is said to have a *bivariate normal* distribution (or to be jointly normally distributed) provided the linear combination $sX + tY$ is normally distributed for each pair $(s,t) \in \mathbf{R}^2$.

3. Notation. Let X and Y have a bivariate normal distribution. Taking $s = 0$ and then $t = 0$ in the Basic Definition, we see that the marginal distributions of X and Y are necessarily normal distributions. In particular, X and Y have moments of all orders. We use the following notation:

$$\mu := E[X], \sigma^2 := \mathrm{Var}(X), \quad \nu := E[Y], \tau^2 := \mathrm{Var}(Y),$$

and write

$$\rho = \rho(X,Y) := \mathrm{Corr}(X,Y) = \mathrm{Cov}(X,Y)/\sigma\tau$$

for the correlation of X and Y. Here $\mathrm{Cov}(X,Y) := E[(X-\mu)(Y-\nu)]$ is the covariance of X and Y.

4. Characteristic Function. In what follows (X,Y) will be a random vector with a bivariate normal distribution, and we shall use the notation of **3**. To avoid trivial cases we assume that $\sigma > 0$ and $\tau > 0$. The (joint) characteristic function of X and Y is defined by

$$\phi_{X,Y}(s,t) := E[\exp(i(sX + tY))], \qquad s,t \in \mathbf{R}.$$

In view of the Basic Definition, $sX + tY \sim \mathcal{N}(s\mu + t\nu, s^2\sigma^2 + 2st\sigma\tau\rho + t^2\tau^2)$, so

$$\phi_{X,Y}(s,t) = \exp\left[i(s\mu + t\nu) - \frac{1}{2}(s^2\sigma^2 + 2st\sigma\tau\rho + t^2\tau^2)\right], \qquad s,t \in \mathbf{R}.$$

5. Independence. Our goal is to compute explicitly the conditional distribution of X given Y in the bivariate normal case. We begin with a warm-up exercise: Compute the conditional expectation $E[X|Y]$. Our calculation is based on the following observation: By the Basic Definition, for any $c \in \mathbf{R}$, the pair $(X - cY, Y)$ has a bivariate normal distribution. The covariance

$$\mathrm{Cov}(X - cY, Y) = \mathrm{Cov}(X,Y) - c\,\mathrm{Cov}(Y,Y) = \sigma\tau\rho - c\tau^2$$

vanishes if and only if $c = c^* := \sigma\rho/\tau$. The random variables $X - c^*Y$ and Y are then independent(!):

$$\begin{aligned}\phi_{X-c^*Y,Y}(s,t) &= \exp\left[i(s\mu - sc^*\nu + t\nu) - \frac{1}{2}(s^2(\sigma^2 - 2c^*\sigma\tau\rho + [c^*]^2\tau^2) + t^2\tau^2))\right] \\ &= \exp\left[is(\mu - sc^*\nu) - \frac{1}{2}s^2(\sigma^2 - 2c^*\sigma\tau\rho + [c^*]^2\tau^2)\right]\exp\left[it\nu - \frac{1}{2}t^2\tau^2\right] \\ &= \phi_{X-c^*Y}(s)\phi_Y(t).\end{aligned}$$

6. Conditional Expectation. In particular,

$$E[X|Y] = E[X - c^*Y|Y] + E[c^*Y|Y] = E[X - c^*Y] + c^*E[Y|Y] = \mu - c^*\nu + c^*Y,$$

where the second equality above follows from the independence of $X - c^*Y$ and Y. We have shown that

$$E[X|Y] = \mu + \frac{\sigma\rho}{\tau}(Y - \nu), \quad \text{a.s.}$$

7. Conditional Distribution, II. I now claim that the conditional distribution F_y of X given $Y = y$ (in the sense of **1** above) is the normal distribution with mean $\mu + c^*(y - \nu)$ and variance $(1-\rho^2)\sigma^2$. To see this let us use $\Phi_y(s) = \exp(is(\mu+c^*(y-\nu)) - \frac{1}{2}s^2(1-\rho^2)\sigma^2)$ to denote the associated characteristic function. Then, using the independence of $X - c^*Y$ and Y for the third equality below:

$$\begin{aligned}
E[\exp(isX)|Y] &= E[\exp(is(X - c^*Y)\exp(isc^*Y)|Y] = E[\exp(is(X - c^*Y)|Y]\exp(isc^*Y) \\
&= E[\exp(is(X - c^*Y)]\exp(isc^*Y) \\
&= \exp\left[is(\mu + c^*(Y - \nu)) - \frac{1}{2}s^2(\sigma^2 - 2c^*\sigma\tau\rho + [c^*]^2\tau^2)\right] \\
&= \exp\left[is(\mu + c^*(Y - \nu)) - \frac{1}{2}s^2(\sigma^2 - 2\frac{\sigma\rho}{\tau}\sigma\tau\rho + \left(\frac{\sigma\rho}{\tau}\right)^2\tau^2)\right] \\
&= \exp\left[is(\mu + c^*(Y - \nu)) - \frac{1}{2}s^2(\sigma^2 - \sigma^2\rho^2)\right] \\
&= \Phi_Y(s)
\end{aligned}$$

It follows that for each $s \in \mathbf{R}$,

$$E[\exp(isX)|Y](\omega) = \Phi_{Y(\omega)}(s) = \int_{\mathbf{R}} e^{isx}F_{Y(\omega)}(dx) \text{ for a.e. } \omega \in \Omega.$$

This confirms that $\{F_y : y \in \mathbf{R}\}$ serves as a regular conditional distribution for X given Y.