MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J — Fall 2008

Lecture 9 — 10/6/2008

CONTINUOUS RANDOM VARIABLES - II

Contents

1. Review of joint distributions
2. Conditional PDFs
3. The bivariate normal distribution
4. Conditional expectation as a random variable
5. Mixed versions of Bayes' rule

1 REVIEW OF JOINT DISTRIBUTIONS

Recall that two random variables X and Y are said to be jointly continuous if there exists a nonnegative measurable function $f_{X,Y}$ such that

$$\mathbb{P}(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\, dv\, du.$$

Once we have in our hands a general definition of integrals, this can be used to establish that for every Borel subset of $\mathbb{R}^2$, we have

$$\mathbb{P}((X,Y) \in B) = \int_B f_{X,Y}(u,v)\, du\, dv.$$

Furthermore, X is itself a continuous random variable, with density f_X given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy.$$

Finally, recall that $\mathbb{E}[g(X)] = \int g(x) f_X(x)\, dx$. Similar to the discrete case, the expectation of $g(X) = X^m$ and $g(X) = (X - \mathbb{E}[X])^m$ is called the mth moment and the mth central moment, respectively, of X. In particular, $\mathrm{var}(X) \triangleq \mathbb{E}[(X - \mathbb{E}[X])^2]$ is the variance of X.

The properties of expectations developed for discrete random variables in Lecture 6 (such as linearity) apply to the continuous case as well. The subsequent development, e.g., for the covariance and correlation, also applies to the continuous case, practically without any changes. The same is true for the Cauchy-Schwarz inequality.

Finally, we note that all of the definitions and formulas have obvious extensions to the case of more than two random variables.

2 CONDITIONAL PDFS

For the case of discrete random variables, the conditional CDF is defined by $F_{X|Y}(x \mid y) = \mathbb{P}(X \le x \mid Y = y)$, for any y such that $\mathbb{P}(Y = y) > 0$. However, this definition cannot be extended to the continuous case because $\mathbb{P}(Y = y) = 0$, for every y. Instead, we should think of $F_{X|Y}(x \mid y)$ as a limit of $\mathbb{P}(X \le x \mid y \le Y \le y + \delta)$, as δ decreases to zero. Note that

$$
\begin{aligned}
F_{X|Y}(x \mid y) &\approx \mathbb{P}(X \le x \mid y \le Y \le y + \delta) \\
&= \frac{\mathbb{P}(X \le x, y \le Y \le y + \delta)}{\mathbb{P}(y \le Y \le y + \delta)} \\
&\approx \frac{\int_{-\infty}^{x} \int_{y}^{y+\delta} f_{X,Y}(u, v)\, dv\, du}{\delta f_Y(y)} \\
&\approx \frac{\delta \int_{-\infty}^{x} f_{X,Y}(u, y)\, du}{\delta f_Y(y)} \\
&= \frac{\int_{-\infty}^{x} f_{X,Y}(u, y)\, du}{f_Y(y)}.
\end{aligned}
$$

The above expression motivates the following definition.

Definition 1. *(a) The conditional CDF of X given Y is defined by*

$$F_{X|Y}(x\,|\,y) = \int_{-\infty}^{x} \frac{f_{X,Y}(u,y)}{f_Y(y)}\,du,$$

for every y such that $f_Y(y) > 0$, where f_Y is the marginal PDF of Y.

(b) The conditional PDF of X given Y is defined by

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

for every y such that $f_Y(y) > 0$.

(c) The conditional expectation of X given $Y = y$ is defined by

$$\mathbb{E}[X\,|\,Y = y] = \int x f_{X|Y}(x\,|\,y)dx,$$

for every y such that $f_Y(y) > 0$.

(d) The conditional probability of the event $\{X \in A\}$, given $Y = y$, is defined by

$$\mathbb{P}(X \in A\,|\,Y = y) = \int_A f_{X|Y}(x\,|\,y)dx,$$

for every y such that $f_Y(y) > 0$.

It can be checked that $F_{X\,|\,Y}$ is indeed a CDF (it satisfies the required properties such as monotonicity, right-continuity, etc.) For example, observe that

$$\lim_{x\to\infty} F_{X|Y}(x\,|\,y) = \int_{-\infty}^{\infty} \frac{f_{X,Y}(u,y)}{f_Y(y)}\,du = 1,$$

since the integral of the numerator is exactly $f_Y(y)$.

Finally, we note that

$$\mathbb{E}[X] = \int \mathbb{E}[X\,|\,Y = y] f_Y(y)\,dy,$$

and

$$\mathbb{P}(X \in A) = \int \mathbb{P}(X \in A\,|\,Y = y) f_Y(y)\,dy.$$

These two relations are established as in the discrete case, by just replacing summations with integrals. They can be rigorously justified if the random variable X is nonnegative or integrable.

3 THE BIVARIATE NORMAL DISTRIBUTION

Let us fix some $\rho \in (-1, 1)$ and consider the function, called the **standard bivariate normal PDF**,

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big(- \frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\Big).$$

Let X and Y be two jointly continuous random variables, defined on the same probability space, whose joint PDF is f.

Proposition 1. *(a) The function f is a indeed a PDF (integrates to 1).*

(b) The marginal density of X and Y is $N(0,1)$, the standard normal PDF.

(c) We have $\rho(X,Y) = \rho$. Also, X and Y are independent iff $\rho = 0$.

(d) The conditional density of X, given $Y = y$, is $N(\rho y, 1-\rho^2)$.

Proof: We will use repeatedly the fact that $1/(\sqrt{2\pi}\sigma)\exp(-\frac{(x-\mu)^2}{2\sigma^2})$ is a PDF (namely, the PDF of the $N(\mu, \sigma^2)$ distribution), and thus integrates to one.

(a)-(b) We note that $x^2 - 2\rho xy + y^2 = x^2 - 2\rho xy + \rho^2 y^2 + (1-\rho^2)y^2$, and obtain

$$\begin{aligned} f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx &= \frac{\exp\Big(-\frac{(1-\rho^2)y^2}{2(1-\rho^2)}\Big)}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\Big(-\frac{(x-\rho y)^2}{2(1-\rho^2)}\Big)dx \\ &= \frac{\exp(-y^2/2)}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\Big(-\frac{(x-\rho y)^2}{2(1-\rho^2)}\Big)dx \end{aligned}$$

But we recognize

$$\frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{\infty} \exp\Big(-\frac{(x-\rho y)^2}{2(1-\rho^2)}\Big)dx$$

as the PDF of the $N(\rho y, 1-\rho^2)$ distribution. Thus, the integral of this density equals one, and we obtain

$$f_Y(y) = \frac{\exp(-y^2/2)}{\sqrt{2\pi}},$$

which is the standard normal PDF. Since $\int_{-\infty}^{\infty} f_Y(y)\,dy = 1$, we conclude that $f(x,y)$ integrates to one, and is a legitimate joint PDF. Furthermore, we have verified that the marginal PDF of Y (and by symmetry, also the marginal PDF of X) is the standard normal PDF, $N(0,1)$.

(c) We have $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY]$, since X and Y are standard normal, and therefore have zero mean. We now have

$$\mathbb{E}[XY] = \int\int xyf(x,y)\,dy\,dx.$$

Applying the same trick as above, we obtain for every y,

$$\int xf(x,y)dx = \frac{\exp(-y^2/2)}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} x\exp\Big(-\frac{(x-\rho y)^2}{2(1-\rho^2)}\Big)dx.$$

But

$$\frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{\infty} x\exp\Big(-\frac{(x-\rho y)^2}{2(1-\rho^2)}\Big)dx = \rho y,$$

since this is the expected value for the $N(\rho y, 1-\rho^2)$ distribution. Thus,

$$\mathbb{E}[XY] = \int\int xyf(x,y)\,dx\,dy = \int y\rho y f_Y(y)\,dy = \rho\int y^2 f_Y(y)\,dy = \rho,$$

since the integral is the second moment of the standard normal, which is equal to one. We have established that $\mathrm{Cov}(X,Y) = \rho$. Since the variances of X and Y are equal to unity, we obtain $\rho(X,Y) = \rho$. If X and Y are independent, then $\rho(X,Y) = 0$, implying that $\rho = 0$. Conversely, if $\rho = 0$, then

$$f(x,y) = \frac{1}{2\pi}\exp\Big(-\frac{x^2+y^2}{2}\Big) = f_X(x)f_Y(y),$$

and therefore X and Y are independent. Note that the condition $\rho(X,Y) = 0$ implies independence, for the special case of the bivariate normal, whereas this implication is not always true, for general random variables.

(d) Let us now compute the conditional PDF. Using the expression for $f_Y(y)$, we have

$$
\begin{aligned}
f_{X|Y}(x\,|\,y) &= \frac{f(x,y)}{f_Y(y)} \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big(-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\Big)\sqrt{2\pi}\exp(y^2/2) \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\Big(-\frac{x^2-2\rho xy+\rho^2y^2}{2(1-\rho^2)}\Big) \\
&= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\Big(-\frac{(x-\rho y)^2}{2(1-\rho^2)}\Big),
\end{aligned}
$$

which we recognize as the $N(\rho y, 1-\rho^2)$ PDF.

$\square$

We have discussed above the special case of a bivariate normal PDF, in which the means are zero and the variances are equal to one. More generally, the bivariate normal PDF is specified by five parameters, $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$, and is given by

$$
f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Big(-\frac{1}{2}Q(x,y)\Big),
$$

where

$$
Q(x,y) = \frac{1}{1-\rho^2}\Big[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)}{\sigma_1}\frac{(y-\mu_2)}{\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\Big].
$$

For this case, it can be verified that

$$
\mathbb{E}[X] = \mu_1, \quad \mathrm{var}(X) = \sigma_1^2, \quad \mathbb{E}[Y] = \mu_2, \quad \mathrm{var}(Y) = \sigma_2^2, \quad \rho(X,Y) = \rho.
$$

These properties can be derived by extending the tedious calculations in the preceding proof.

There is a further generalization to more than two random variables, resulting in the multivariate normal distribution. It will be carried out in a more elegant manner in a later lecture.

4 CONDITIONAL EXPECTATION AS A RANDOM VARIABLE

Similar to the discrete case, we define $\mathbb{E}[X\,|\,Y]$ as a random variable that takes the value $\mathbb{E}[X\,|\,Y=y]$, whenever $Y=y$, where $f_Y(y)>0$. Formally, $\mathbb{E}[X\,|\,Y]$ is a function $\psi:\Omega\to\mathbb{R}$ that satisfies

$$
\psi(\omega) = \int x f_{X|Y}(x\,|\,y)\,dx,
$$

for every ω such that $Y(\omega) = y$, where $f_Y(y) > 0$. Note that nothing is said about the value of $\psi(\omega)$ for those ω that result in a y at which $f_Y(y) = 0$. However, the set of such ω has zero probability measure. Because, the value of $\psi(\omega)$ is completely determined by the value of $Y(\omega)$, we also have $\psi(\omega) = \phi(Y(\omega))$, for some function $\phi : \mathbb{R} \to \mathbb{R}$. It turns out that both functions ψ and ϕ can be taken to be measurable.

One might expect that when X and Y are jointly continuous, then $\mathbb{E}[X \mid Y]$ is a continuous random variable, but this is not the case. To see this, suppose that X and Y are independent, in which case $\mathbb{E}[X \mid Y = y] = \mathbb{E}[X]$, which also implies that $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$. Thus, $\mathbb{E}[X \mid Y]$ takes a constant value, and is therefore a trivial case of a discrete random variable.

Similar to the discrete case, for every measurable function g, we have

$$\mathbb{E}\big[\mathbb{E}[X \mid Y] g(Y)\big] = \mathbb{E}[X g(Y)] \tag{1}$$

(assuming all expectations involved to be well-defined). The proof is essentially the same, with integrals replacing summations. In fact, this property can be taken as a more abstract definition of the conditional expectation. By letting g be identically equal to 1, we obtain

$$\mathbb{E}\big[\mathbb{E}[X \mid Y]\big] = \mathbb{E}[X].$$

Example. We have a stick of unit length $[0,1]$, and break it at X, where X is uniformly distributed on $[0,1]$. Given the value x of X, we let Y be uniformly distributed on $[0,x]$, and let Z be uniformly distributed on $[0,1-x]$. We assume that conditioned on $X = x$, the random variables Y and Z are independent. We are interested in the distribution of Y and Z, their expected values, and the expected value of their product.

It is clear from symmetry that Y and Z have the same marginal distribution, so we focus on Y. Let us first find the joint distribution of Y and X. We have $f_X(x) = 1$, for $x \in [0,1]$, and $f_{Y|X}(y \mid x) = 1/x$, for $y \in [0,x]$. Thus, the joint PDF is

$$f_{X,Y}(x,y) = f_{Y|X}(y \mid x) f_X(x) = \frac{1}{x} \cdot 1 = \frac{1}{x}, \qquad 0 \le y \le x \le 1.$$

We can now find the PDF of Y:

$$f_Y(y) = \int_y^1 f_{X,Y}(x,y)\,dx = \int_y^1 \frac{1}{x}\,dx = \log x\Big|_y^1 = \log(1/y).$$

(check that this indeed integrates to unity). Integrating by parts, we then obtain

$$\mathbb{E}[Y] = \int_0^1 y f_Y(y) dy = \int_0^1 y \log(1/y)\,dy = \frac{1}{4}.$$

The above calculation is more involved than necessary. For a simpler argument, simply observe that $\mathbb{E}[Y \mid X = x] = x/2$, since Y conditioned on $X = x$ is uniform on $[0, x]$. In particular, $\mathbb{E}[Y \mid X] = X/2$. It follows that $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[X/2] = 1/4$.

For an alternative version of this argument, consider the random variable Y/X. Conditioned on the event $X = x$, this random variable takes values in the range $[0, 1]$, is uniformly distributed on that range, and has mean $1/2$. Thus, the conditional PDF of Y/X is not affected by the value x of X. This implies that Y/X is independent of X, and we have

$$\mathbb{E}[Y] = \mathbb{E}[(Y/X)X] = \mathbb{E}[Y/X] \cdot \mathbb{E}[X] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

To find $\mathbb{E}[YZ]$ we use the fact that, conditional on $X = x$, Y and Z are independent, and obtain

$$\begin{aligned}\mathbb{E}[YZ] &= \mathbb{E}\big[\mathbb{E}[YZ \mid X]\big] = \mathbb{E}\big[\mathbb{E}[Y \mid X] \cdot \mathbb{E}[Z \mid X]\big] \\ &= \mathbb{E}\Big[\frac{X}{2} \cdot \frac{1-X}{2}\Big] = \int_0^1 \frac{x(1-x)}{4}\, dx = \frac{1}{24}.\end{aligned}$$

Exercise 1. Find the joint PDF of Y and Z. Find the probability $\mathbb{P}(Y + Z \leq 1/3)$. Find $\mathbb{E}[X|Y]$, $\mathbb{E}[X|Z]$, and $\rho(Y, Z)$.

4.1 Optimality properties of conditional expectations

The conditional expectation $\mathbb{E}[X \mid Y]$ can be viewed as an estimate of X, based on the value of Y. In fact, it is an optimal estimate, in the sense that the mean square of the resulting estimation error, $X - \mathbb{E}[X \mid Y]$, is as small as possible.

Theorem 1. *Suppose that $\mathbb{E}[X^2] < \infty$. Then, for any measurable function $g : \mathbb{R} \to \mathbb{R}$, we have*

$$\mathbb{E}\big[(X - \mathbb{E}[X \mid Y])^2\big] \leq \mathbb{E}\big[(X - g(Y))^2\big].$$

Proof: We have

$$\begin{aligned}\mathbb{E}\big[(X - g(Y))^2\big] &= \mathbb{E}[(X - \mathbb{E}[X \mid Y])^2\big] + \mathbb{E}\big[(\mathbb{E}[X \mid Y] - g(Y))^2\big] \\ &\quad + \mathbb{E}[(X - \mathbb{E}[X \mid Y])(\mathbb{E}[X \mid Y] - g(Y))\big] \\ &\geq \mathbb{E}[(X - \mathbb{E}[X \mid Y])^2\big].\end{aligned}$$

The inequality above is obtained by noticing that the term $\mathbb{E}\big[(X - g(Y))^2\big]$ is always nonnegative, and that the term $\mathbb{E}[(X - \mathbb{E}[X \mid Y])(\mathbb{E}[X \mid Y] - g(Y))\big]$

is of the form $\mathbb{E}\big[(X - \mathbb{E}[X \mid Y])\psi(Y)\big]$ for $\psi(Y) = \mathbb{E}[X \mid Y] - g(Y)$, and is therefore equal to zero, by Eq. (1). $\square$

Notice that the preceding proof only relies on the property (1). As we have discussed, we can view this as the defining property of conditional expectations, for general random variables. It follows that the preceding theorem is true for all kinds of random variables.

5 MIXED VERSIONS OF BAYES' RULE

Let X be an unobserved random variable, with known CDF, F_X. We observe the value of a related random variable, Y, whose distribution depends on the value of X. This dependence can be captured by a conditional CDF, $F_{Y|X}$. On the basis of the observed value y of Y, would like to make an inference on the unknown value of X. While sometimes, this inference aims at a numerical estimate for X, the most complete answer, which includes everything that can be said about X, is the conditional distribution of X, given Y. This conditional distribution can be obtained by using an appropriate form of Bayes' rule.

When X and Y are both discrete, Bayes' rule takes the simple form

$$p_{X|Y}(x \mid y) = \frac{p_X(x)p_{Y|X}(y \mid x)}{p_Y(y)} = \frac{p_X(x)p_{Y|X}(y \mid x)}{\sum_{x'} p_X(x')p_{Y|X}(y \mid x')}.$$

When X and Y are both continuous, Bayes' rule takes a similar form,

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

which follows readily from the definition of the conditional PDF.

It remains to consider the case where one random variable is discrete and the other continuous. Suppose that K is a discrete random variable and Z is a continuous random variable. We describe their joint distribution in terms of a function $f_{K,Z}(k, z)$ that satisfies

$$\mathbb{P}(K = k, Z \leq z) = \int_{-\infty}^{z} f_{K,Z}(k, t)\, dt.$$

We then have

$$p_K(k) = \mathbb{P}(K = k) = \int_{-\infty}^{\infty} f_{K,Z}(k, t)\, dt,$$

and[1]

$$F_Z(z) = \mathbb{P}(Z \le z) = \sum_k \int_{-\infty}^{z} f_{K,Z}(k,t)\,dz = \int_{-\infty}^{z} \sum_k f_{K,Z}(k,t)\,dz,$$

which implies that

$$f_Z(z) = \sum_k f_{K,Z}(k,z),$$

is the PDF of Z.

Note that if $\mathbb{P}(K = k) > 0$, then

$$\mathbb{P}(Z \le z \mid K = k) = \int_{-\infty}^{z} \frac{f_{K,Z}(k,t)}{p_K(k)}\,dt,$$

and therefore, it is reasonable to define

$$f_{Z|K}(z \mid k) = f_{K,Z}(k,z)/p_K(k).$$

Finally, for z such that $f_Z(z) > 0$, we define $p_{K|Z}(k \mid z) = f_{K,Z}(k,z)/f_Z(z)$, and interpret it as the conditional probability of the event $K = k$, given that $Z = z$. (Note that we are conditioning on a zero probability event; a more accurate interpretation is obtained by conditioning on the event $z \le Z \le z + \delta$, and let $\delta \to 0$.) With these definitions, we have

$$f_{K,Z}(k,z) = p_K(k) f_{Z|K}(z \mid k) = f_Z(z) p_{K|Z}(k \mid z),$$

for every (k,z) for which $f_{K,Z}(k,z) > 0$. By rearranging, we obtain two more versions of the Bayes' rule:

$$f_{Z|K}(z \mid k) = \frac{f_Z(z) p_{K|Z}(k \mid z)}{p_K(k)} = \frac{f_Z(z) p_{K|Z}(k \mid z)}{\int f_Z(z') p_{K|Z}(k \mid z')\,dz'},$$

and

$$p_{K|Z}(k \mid z) = \frac{p_K(k) f_{Z|K}(z \mid k)}{f_Z(z)} = \frac{p_K(k) f_{Z|k}(z \mid k)}{\sum_{k'} p_K(k') f_{Z|K}(z \mid k')}.$$

Note that all four versions of Bayes' rule take the exact same form; the only difference is that we use PMFs and summations for discrete random variables, as opposed to PDFs and integrals for continuous random variables.

[1]The interchange of the summation and the integration can be rigorously justified, because the terms inside are nonnegative.

6.436J / 15.085J Fundamentals of Probability
Fall 2008