Math 425
Intro to Probability
Lecture 35

Kenneth Harris
kaharri@umich.edu

Department of Mathematics
University of Michigan

April 13, 2009

Definition – Continuous Case

☞ Let X and Y be jointly continuous random variables with joint density $f(x, y)$. The conditional probability density function of X given that $Y = y$ was defined for all y with $f_Y(y) > 0$ by

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

☞ Since $f_{X|Y}(x|y)$ is a density, we can define the conditional expectation of X given $Y = y$, for all values y with $f_Y(y) > 0$ by

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

Example– Conditional density

Example. Compute the conditional densities and expectation for the following joint density of X and Y.

$$f(x, y) = x(y - x)e^{-y} \qquad 0 \le x \le y < \infty.$$

☞ The marginal densities are

$$f_Y(y) = \int_0^y x(y - x)e^{-y}\, dx = \frac{y^3}{6}e^{-y}$$

$$f_X(x) = \int_x^{\infty} x(y - x)e^{-y}\, dy = xe^{-x}$$

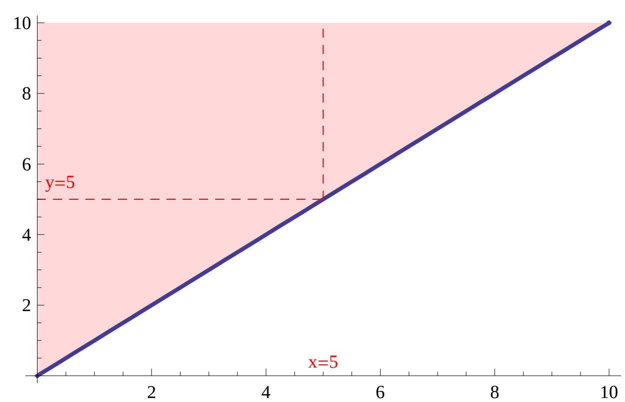
So, the conditional densities are

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{x(y - x)e^{-y}}{xe^{-x}} = (y - x)e^{-(y-x)}$$

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(x)} = \frac{x(y - x)e^{-y}}{\frac{y^3}{6}e^{-y}} = 6x(y - x)y^{-3}$$

Region of Integration

☞ Region of integration: $0 \le x \le y < \infty$.



Example – Conditional Expectation

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = (y-x)e^{-(y-x)}$$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(x)} = 6x(y-x)y^{-3}$$

☞ Compute the conditional expectations.

$$E[X|Y=y] = \int_0^y x \cdot \left(6x(y-x)y^{-3}\right) dx = \frac{1}{2}y$$

$$E[Y|X=x] = \int_x^\infty y \cdot \left((y-x)e^{-(y-x)}\right) dy = x+2$$

☞ Note the following connections

When $Y = y$ $\quad E[X|Y=y] = \frac{1}{2}y$

When $X = x$ $\quad E[Y|X=x] = x+2$

Conditional Expectation Function

☞ We write $E[X|Y]$ for the function of the variable Y given by

$$Y = y \xrightarrow{E[X|Y]} E[X|Y=y].$$

In the previous example

$$E[X|Y] = \frac{1}{2}Y$$
$$E[Y|X] = X+2$$

We can take expectations of $E[X|Y]$:

$$E\big[E[X|Y]\big] = \sum_y E[X|Y=y] \cdot p_Y(y) \quad \text{when } X, Y \text{ discrete}$$

$$E\big[E[X|Y]\big] = \int_{-\infty}^{\infty} E[X|Y=y] \cdot f_Y(y) \quad \text{when } X, Y \text{ continuous}$$

Conditioning

☞ We stated the following in the discrete case (using different terminology).

Proposition (Conditioning)

$$E[X] = E\big[E[X|Y]\big].$$

Equivalently, (spelling this out)

$$E[X] = \sum_y E[X|Y=y] \cdot p_Y(y) \quad \textit{when } X, Y \textit{ discrete}$$

$$E[X] = \int_{-\infty}^{\infty} E[X|Y=y] \cdot f_Y(y) \quad \textit{when } X, Y \textit{ continuous}$$

Proof of Proposition

Proof.

We give a proof for continuous X and Y. Recall, for a function of Y, $h(Y)$:

$$E[h(Y)] = \int_{-\infty}^{\infty} h(y) \cdot f_Y(y)\, dy$$

Then

$$\begin{aligned}
E\big[E[X|Y]\big] &= \int_{-\infty}^{\infty} E[X|Y=y] \cdot f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\, dx\right) f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y) f_Y(y)\, dy\, dx \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy\right) dx \\
&= \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx = E[X].
\end{aligned}$$

□

Example – continued

Example.
Let X and Y have the joint density (previous example)

$$f(x,y) = x(y-x)e^{-y} \qquad 0 \le x \le y < \infty.$$

Compute the expectations $E[X]$ and $E[Y]$.

Solution. From the previous example

$$\begin{aligned} E[X] &= E[E[X|Y]] = \int_0^\infty \frac{y}{2} \cdot f_Y(y)\,dy \\ &= \int_0^\infty \frac{y}{2} \cdot \frac{y^3}{6} e^{-y} = 2 \\ E[Y] &= E[E[Y|X]] = \int_0^\infty (x+2) \cdot f_X(x)\,dx \\ &= \int_0^\infty (x+2) \cdot xe^{-x}\,dx = 4 \end{aligned}$$

Mixed Probability

☞ We have dealt with jointly continuous and jointly discrete random variables. However, it still makes sense to talk about the joint distribution of a continuous and a discrete random variable.

☞ Let N be a discrete random variable and n a value with $\mathbf{P}\{N = n\} > 0$. Conditioning on $N = n$, $\mathbf{P}(\cdot \mid N = n)$, is a probability.

☞ If Y is a continuous random variable, then $F_{Y|N}(\cdot|n)$ is the cumulative distribution of Y given $N = n$:

$$F_{Y|N}(y|n) = \mathbf{P}(Y \le y \mid N = n).$$

☞ We also have a conditional density of Y given $N = n$, $f_{Y|N}(y|n)$:

$$\begin{aligned} f_{Y|N}(y) &= \frac{d}{dy} F_{Y|N}(y) \\ &= \lim_{\Delta y \to 0} \frac{\mathbf{P}(y \le Y \le y + \Delta y \mid A)}{\Delta y} \end{aligned}$$

Mixed conditional density

☞ It is also possible to condition a discrete random variable N given a continuous random variable $Y = y$, where $f_Y(y) > 0$.
We must be careful since $\mathbf{P}\{Y = y\} = 0$ for all y.

☞ Use Bayes theorem on a small interval around $Y = y$:

$$\begin{aligned} \mathbf{P}(N = n \mid y \le Y \le y + \Delta y) &= \frac{\mathbf{P}(y \le Y \le y + \Delta y \mid N = n)}{\mathbf{P}\{y \le Y \le y + \Delta y\}} \cdot \mathbf{P}\{N = n\} \\ &= \frac{F_{Y|N}(y + \Delta y|n) - F_{Y|N}(y|n)}{F_Y(y + \Delta y) - F_Y(y)} \cdot \mathbf{P}\{N = n\} \end{aligned}$$

☞ Compute the limit as $\Delta y \to 0$ (just as we did with joint density).

Mixed conditional density

We now define the probability $N = n$ given $Y = y$:

$$\begin{aligned} \mathbf{P}(N = n \mid Y = y) &= \lim_{\Delta y \to 0} \mathbf{P}(N = n \mid y \le Y \le y + \Delta y) \\ &= \lim_{\Delta y \to 0} \frac{F_{Y|N}(y + \Delta y|n) - F_{Y|N}(y|n)}{F_Y(y + \Delta y) - F_Y(y)} \cdot \mathbf{P}\{N = n\} \\ &= \lim_{\Delta y \to 0} \frac{\frac{F_{Y|N}(y+\Delta y|n) - F_{Y|N}(y|n)}{\Delta y}}{\frac{F_Y(y+\Delta y) - F_Y(y)}{\Delta y}} \cdot \mathbf{P}\{N = n\} \\ &= \frac{\frac{d}{dy} F_{Y|N}(y|n)}{\frac{d}{dy} F_Y(y)} \cdot \mathbf{P}\{N = n\} \\ &= \frac{f_{Y|N}(y|n)}{f_Y(y)} \cdot \mathbf{P}\{N = n\}. \end{aligned}$$

Computing Probabilities by Conditioning

☞ Let A be some event, and let I_A be the indicator variable

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

☞ When Y is continuous the conditional probability of A given $Y = y$ (provided $f_Y(y) > 0$) is

$$\mathbf{P}(A \mid Y = y) \;=\; \mathbf{P}(I_A = 1 \mid Y = y) \;=\; E[I_A | Y = y].$$

☞ We can compute $\mathbf{P}\{A\}$ by conditioning on Y:

$$\begin{aligned} \mathbf{P}\{A\} &= E[I_A] = E\big[\, E[I_A | Y] \,\big] \\ &= \int_{-\infty}^{\infty} E[I_A | Y = y] \cdot f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} \mathbf{P}(A \mid Y = y) \cdot f_Y(y)\, dy \end{aligned}$$

Example

☞ We can use conditioning to simply compute probabilities that could have been computed by Chapter 6 methods (although often not as simply).

See Examples 7.5l and 7.5m in Ross.

Example/Theorem. If X and Y are independent continuous random variables, then

$$\mathbf{P}\{X < Y\} \;=\; \int_{-\infty}^{\infty} \mathbf{P}\{X < y\} \cdot f_Y(y)\, dy$$

where for any real number y,

$$\mathbf{P}\{X < y\} \;=\; \int_{-\infty}^{y} f_X(x)\, dx.$$

Proof

☞ Condition on the value of Y:

$$\begin{aligned} \mathbf{P}\{X < Y\} &= \int_{-\infty}^{\infty} \mathbf{P}(X < Y \mid Y = y) \cdot f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} \mathbf{P}(X < y \mid Y = y) \cdot f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} \mathbf{P}\{X < y\} \cdot f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{y} f_X(x)\, dx \right) \cdot f_Y(y)\, dy \end{aligned}$$

The next-to-last line is because X and Y are independent.

Example

Example. Let X be exponential with parameter λ and Y be exponential with parameter μ. Suppose that X and Y are independent. Then

$$\begin{aligned} \mathbf{P}\{X < Y\} &= \int_{-\infty}^{\infty} \mathbf{P}\{X < y\} \cdot f_Y(y)\, dy \\ &= \int_{0}^{\infty} \left(1 - e^{-\lambda y}\right) \cdot \mu e^{-\mu y}\, dy \\ &= \frac{\lambda}{\mu + \lambda}. \end{aligned}$$

Example

Example

A (very) large urn contains a number of coins, and each of these coins has some probability p of turning up heads when flipped. Suppose that the composition of the coins in the urn is such that when a coin is chosen at random, then its p-value can be regarded as uniformly distributed over $[0, 1]$.

(a) If a coin is chosen at random and flipped twice, what is the probability that both flips are heads?
(b) If a coin is chosen at random and flipped three times, what is the probability that exactly two flips are heads?
(c) What is the expected number of flips of a randomly chosen coins before the first heads appears?

Example – part (a)

☞ Let A be the event that both flips are heads and Y the random variable giving the probability of heads for a randomly chosen coin.

☞ Condition on $Y = p$:

$$\begin{aligned} \mathbf{P}\{A\} &= \int_0^1 \mathbf{P}(A \mid Y = p)\, dp \\ &= \int_0^1 p^2\, dp \\ &= \frac{1}{3}. \end{aligned}$$

The probability of getting two heads from a randomly chosen coin is a bit better than two heads from a fair coin, $\frac{1}{4}$.

Example – part (b)

☞ What is the probability of getting exactly two heads in three flips? Let this event be B and condition on $Y = p$:

$$\begin{aligned} \mathbf{P}\{B\} &= \int_0^1 \mathbf{P}(B \mid Y = p)\, dp \\ &= \int_0^1 \binom{3}{2} p^2(1-p)\, dp \\ &= \binom{3}{2} \cdot \int_0^1 p^2(1-p)\, dp \\ &= 3 \cdot \frac{1}{12} \\ &= \frac{1}{4}. \end{aligned}$$

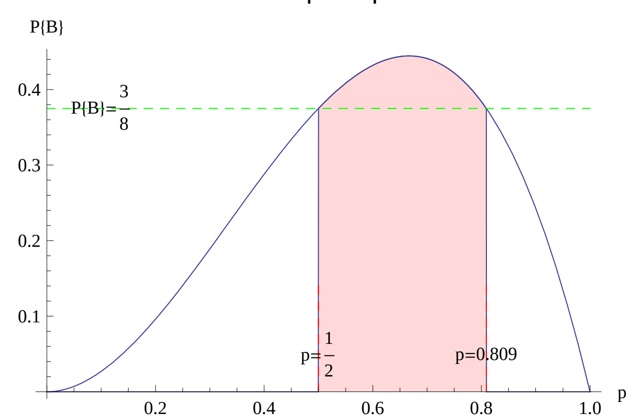
The probability of getting exactly two heads in three flips from a randomly chosen coin is a bit poorer than from a fair coin, $\frac{3}{8}$.

Example – part (b)

☞ Plot of $3(p^2 - p^3)$. This the density function for the likelihood of a randomly chosen coin to produce exactly two heads in three flips.

The region where the random coin does better than a fair coin is

$$0.5 \le p \le \frac{1}{4} + \frac{\sqrt{5}}{4} \approx 0.809$$



Example – part (c)

☞ How many flips can be expected before a heads appears on a randomly chosen coin?

Let N be the number of flips required until a heads appears for a randomly chosen coin. Condition on Y:

$$
\begin{aligned}
\mathbf{P}\{N=i\} &= \int_0^1 \mathbf{P}(N=i \mid Y=p)\,dp \\
&= \int_0^1 (1-p)^{i-1} p\,dp \\
&= \int_0^1 q^{i-1}(1-q)\,dq \qquad q = 1-p \\
&= \frac{1}{i} - \frac{1}{i+1} \\
&= \frac{1}{i(i+1)}.
\end{aligned}
$$

Example – part (c)

☞ The expected number of flips until a heads appears from a randomly chosen coin

$$
\begin{aligned}
E[N] &= \sum_{i=1}^{\infty} i \cdot \mathbf{P}\{N=i\} \\
&= \sum_{i=1}^{\infty} i \cdot \frac{1}{i(i+1)} \\
&= \sum_{i=1}^{\infty} \frac{1}{i+1} \\
&= \infty
\end{aligned}
$$

Example

Example

Fix n and let N be the binomial random variable giving the number of heads in n tosses of a randomly chosen coin from our urn.
Compute the probability mass function for N.

☞ Condition on Y.

$$
\begin{aligned}
\mathbf{P}\{N=k\} &= \int_0^1 \mathbf{P}(N=k \mid Y=p)\,dp \\
&= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k}\,dp \\
&= \binom{n}{k} \int_0^1 p^k (1-p)^{n-k}\,dp
\end{aligned}
$$

Example – continued

☞ Define $B(a,b)$ for positive integers a and b

$$
B(a,b) = \int_0^1 p^a (1-p)^b\,dp
$$

☞ Use integration by parts with $u = (1-p)^b$ and $dv = p^a$:

$$
\begin{aligned}
\int_0^1 p^a(1-p)^b\,dp &= \frac{b}{a+1} \int_0^1 p^{a+1}(1-p)^{b-1}\,dp \\
&= \frac{b}{a+1} \cdot B(a+1, b-1)
\end{aligned}
$$

☞ So,

$$
B(a,b) = \frac{b}{a+1} \cdot B(a+1, b-1)
$$

Example – continued

$$B(a,b) = \frac{b}{a+1} \cdot B(a+1, b-1)$$

☞ Carry this out b time:

$$\begin{aligned} B(a,b) &= \frac{b!}{(a+1)\cdots(a+b)} \cdot B(a+b, 0) \\ &= \frac{b!}{(a+1)\cdots(a+b)} \cdot \int_0^1 p^{a+b}\,dp \\ &= \frac{b!}{(a+1)\cdots(a+b)\cdot(a+b+1)} \\ &= \frac{b!\cdot a!}{(a+b+1)!} \end{aligned}$$

☞ So,

$$B(a,b) = \int_0^1 p^a(1-p)^b\,dp = \frac{b!\cdot a!}{(a+b+1)!}$$

Example – continued

☞ Back to our original problem:

$$\begin{aligned} \mathbf{P}\{N=k\} &= \binom{n}{k}\int_0^1 p^k(1-p)^{n-k}\,dp \\ &= \binom{n}{k}\cdot B(k, n-k) \\ &= \binom{n}{k}\cdot\frac{k!\cdot(n-k)!}{(n+1)!} \\ &= \frac{1}{n+1} \end{aligned}$$

☞ If we choose a coin at random and flip it n times, then the number of heads is uniformly distributed among the possible values $0, 1, \ldots, n$.

Beta Distribution

☞ Let's compute the conditional density of Y (the bias of the coin) given k heads were observed (so $N=k$), $f_{Y|N}(p|n)$. Recall :

$$\mathbf{P}(N=k\mid Y=p) = \frac{f_{Y|N}(p|k)}{f_Y(p)}\cdot PR\{N=k\}$$

☞ Solving for $f_{Y|N}(p|k)$:

$$\begin{aligned} f_{Y|N}(p|k) &= \frac{\mathbf{P}(N=k\mid Y=p)\cdot f_Y(p)}{\mathbf{P}\{N=k\}} \\ &= \frac{\binom{n}{k}\cdot p^k(1-p)^{n-k}}{\frac{1}{n+1}} \\ &= \frac{(n+1)!}{(n-k)!k!}\cdot p^k(1-p)^{n-k}. \end{aligned}$$

☞ This density is important enough to have a name: a beta density.

Beta Distribution

☞ The beta density function with parameters a and b (where a and b are nonnegative integers) is

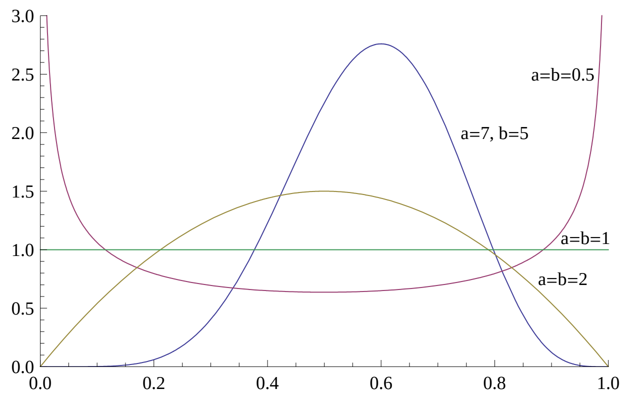
$$f(x) = \begin{cases} \frac{(a+b-1)!}{(a-1)!(b-1)!}\cdot x^{a-1}(1-x)^{b-1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

☞ The conditional density function giving the likelihood of the bias of a coin which is uniformly distributed in $(0,1)$, given the number of heads when flipped n times is beta distributed with parameters $n+1$ and $n-k+1$.

See Ross Section 5.6.4 for the most general form of the beta distribution, using the gamma function, which is the reason for subtracting 1 from a and b. We will only use this density for positive integer values of a and b.

Plot of beta density

☞ Plot of beta density function for several values of the parameters *a* and *b*.



Beta Distribution

Example. A randomly chosen coin is tossed 10 times and comes up heads 6 times.

What is the likelihood that the bias of the coin to heads is between 0.3 and 0.7?

☞ The prior probability of $Y \in [0.3, 0.7]$ (before we observed the sequence of 10 flips) is

$$\mathbf{P}\{0.3 \leq Y \leq 0.7\} = 0.4.$$

☞ Once we observe the 10 flips and see 6 coins, we use the beta density $f_{Y|N}(p|n)$:

$$\begin{aligned} \mathbf{P}(0.3 \leq Y \leq 0.7 \,|\, X = 6) &= \int_{0.3}^{0.7} \frac{(a+b+1)!}{a!b!} \cdot p^k (1-p)^{n-k} \, dp \\ &\approx 0.768 \end{aligned}$$

I computed this value using Mathematica's built-in beta distribution.