**Stat210A: Theoretical Statistics** **Recitation Date: September 8, 2014**

# Recitation 2: Conditional Distributions

*Lecturer: Tamara Broderick* *Scribe: Johnny Hong*

## 1 Miscellany/Review

### 1.1 Convergence in Probability

(Notation p. XI or p. 129 in Sec. 8.1 in Keener)

**Definition 1.** A sequence of random variables $(Y_n)$ *converges in probability* to a random variable $Y$ as $n \to \infty$ ($Y_n \xrightarrow{P} Y$) if for all $\epsilon > 0$, $P(|Y_n - Y| \geq \epsilon) \to 0$ as $n \to \infty$.

### 1.2 Density

**Definition 2.** A measure $\nu$ has *density* $f$ with respect to a measure $\mu$ if

$$\nu(A) = \int_A f(x)\mu(dx) = \int_{x \in \mathcal{X}} f(x) 1_A(x) \mu(dx). \tag{1}$$

If $P_X \ll$ counting measure on $\mathbb{Z}^D$,

$$P_X(A) = P(X \in A) = \sum_{x \in \mathbb{Z}^D} p(x) 1_A(x) \tag{2}$$

This is the form of density in the discrete case.

If $P_X \ll$ Lebesgue measure on $\mathbb{R}^D$,

$$P_X(A) = \int_{x \in \mathbb{R}^D} p(x) 1_A(x) dx. \tag{3}$$

## 2 Joint Density

Consider $Z = (X, Y) \in \mathbb{R}^{m+n}$, where $X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^n$. Then

$$P_Z(A) = \int p_Z(z) 1_A(z) \mu_Z(dz). \tag{4}$$

**Definition 3.** Suppose $\mu_z(dz) = (\mu \times \nu)(dx, dy) = \mu(dx)\nu(dy)$ and suppose

$$P_Z(A) = \int p_Z(x, y) 1_A(x, y) \mu(dx) \nu(dy). \tag{5}$$

Then $p_Z$ is the *joint density* of $X$ and $Y$.

# 3 Marginal Density

**Definition 4.** Suppose $X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^n$, and suppose $P_Z \ll \mu \times \nu$ ($\mu$ on $\mathbb{R}^m, \nu$ on $\mathbb{R}^n$). Let $P_X(A) = P(X \in A)$. The density of $P_X$ with respect to $\mu$ is called the *marginal density* of $X$.

**Remark:** This definition implicitly assumes the existence of the density of $P_X$. To show the existence of such density, note that[1]

$$\begin{aligned} \mu(A) = 0 &\implies (\mu \times \nu)(A \times \mathbb{R}^n) = 0 \text{ (by definition of product measure)} \\ &\implies P_Z(A \times \mathbb{R}^n) = 0 \\ &\implies P_X(A) = 0 \quad (P_Z(A \times \mathbb{R}^n) = P_X(A)) \end{aligned} \tag{6}$$

Then Radon-Nikodym Theorem tells us that the density of $X$ exists, so the definition makes sense!

More explicitly,

$$\begin{aligned} P_X(A) &= P(X \in A) \\ &= P(Z \in A \times \mathbb{R}^n) \\ &= \int 1_{A \times \mathbb{R}^n}(x, y) p_Z(x, y) \mu(dx) \nu(dy) \\ &= \int_{\mathbb{R}^m} 1_A(x) \left[ \int_{\mathbb{R}^n} p_Z(x, y) \nu(dy) \right] \mu(dx). \end{aligned} \tag{7}$$

By definition, we have

$$p_X(x) = \int_{\mathbb{R}^n} p_Z(x, y) \nu(dy). \tag{8}$$

**Example 6.1** Let $\mu$ be the counting measure on $\mathbb{Z}$ and $\nu$ be the Lebesgue measure on $\mathbb{R}$. Consider

$$p_Z(x, y) = \binom{k}{x} y^x (1-y)^{k-x} 1_{\{x \in \{0,1,\ldots,k\}, y \in (0,1)\}} \tag{9}$$

Then[2]

$$\begin{aligned} p_X(x) &= \int_0^1 p_Z(x, y) dy \\ &= \binom{k}{x} \frac{\Gamma(x+1)\Gamma(k-x+1)}{\Gamma(k+2)} 1_{\{x \in \{0,1,\ldots,k\}\}} \\ &= \frac{k!}{x!(k-x)!} \frac{x!(k-x)!}{(k+1)!} 1_{\{x \in \{0,1,\ldots,k\}\}} \\ &= \frac{1}{k+1} 1_{\{x \in \{0,1,\ldots,k\}\}} \end{aligned} \tag{11}$$

[1] We can assume $\nu$ is finite without loss of generality.

[2] Recall that $W \sim \text{Beta}(\alpha, \beta)$ if $p_W(w) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} w^{\alpha-1}(1-w)^{\beta-1} 1_{\{w \in (0,1)\}}$, where $\Gamma$ is the gamma function (see Ex. 26 in Ch. 1 in Keener). Some basic properties of the gamma function are as follows:

$$\begin{aligned} &\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) \text{ for } \alpha - 1 > 0 \\ &\Gamma(n) = (n-1)! \text{ for } n - 1 > 0 \\ &\Gamma(1) = 1 \end{aligned} \tag{10}$$

and[3]

$$p_Y(y) = \sum_{x=0}^{k} p_Z(x,y) = \sum_{x=0}^{k} \binom{k}{x} y^x (1-y)^{k-x} 1_{\{y\in(0,1)\}} = 1_{\{y\in(0,1)\}}. \tag{12}$$

Hence, $X \sim \text{Uniform}\{0, 1, ..., k\}$ and $Y \sim \text{Uniform}(0, 1)$.

# 4 Conditional Stuff

**Definition 5.** Let Y, X be two random vectors. Then $Q$ is a *conditional distribution* for $Y$ given $X$ ($Y|(X = x) \sim Q_x$) if

1. For all $x$, $Q_x(\cdot)$ is a probability measure.
2. For all $B$, $Q_x(B)$ is a function of $x$.
3. For all $A$ and $B$, $P(X \in A, Y \in B) = \int_A Q_x(B) P_X(dx)$.

**Intuition: Discrete Case**

$$Q_x(B) = P(Y \in B|X = x) = \frac{P(Y \in B, X = x)}{P(X = x)}. \tag{13}$$

We want to check that this proposed conditional distribution obeys Definition 5.

1. Check that $Q_x(\cdot)$ is a measure. Sum $Q_x(\{y\})$ over all possible values of $y$, and check equals 1.
2. Check that $Q_\cdot(B)$ is a function of $x$. Aside for intuition: $E(Y|X) = \sum_y yP(Y = y|X = x)$
3. $P(X \in A, Y \in B) = \sum_{x\in A} \sum_{y\in B} P(X = x, Y = y) = \sum_{x\in A} \sum_{y\in B} P(Y = y|X = x)P(X = x) = \int_A Q_x(B)P_X(dx)$.

**Corollary 6.** *Let $X, Y$ have joint density $p_Z(x, y)$ with respect to $\mu \times \nu$. Then the conditional distribution $Q_x$ has a density*

$$p_{Y|X}(y|x) = \begin{cases} p_0(y) & \text{if } p_X(x) = 0 \\ \frac{p_Z(x,y)}{p_X(x)} & \text{if } p_X(x) > 0 \end{cases}$$

*where $p_0(y)$ is any fixed density.*

**Remark:** In introductory probability courses, one learns that

$$\text{Conditional Density} = \frac{\text{Joint Density}}{\text{Marginal Density}} \tag{14}$$

*Proof.* Check Definition 5.

[3] Recall that $\binom{k}{x} y^x (1-y)^{k-x} = \text{Binom}(x|k, y)$

1. Check that $Q_x(\cdot)$ is a measure. If $p_X(x) > 0$, notice $\int \frac{p_Z(x,y)}{p_X(x)} dy = \frac{1}{p_X(x)} \int p_Z(x,y) dy = \frac{1}{p_X(x)} p_X(x) = 1$.
2. Check that $Q_\cdot(B)$ is a function of $x$.
3. Leave it as an exercise.

$\square$

**Example continued** Recall that

$$p_Z(x,y) = \binom{k}{x} y^x (1-y)^{k-x} 1_{\{x \in \{0,1,\ldots,k\}, y \in (0,1)\}} \tag{15}$$

We found that $p_Y(y) = 1_{\{y \in (0,1)\}}$. Therefore,

$$p_{X|Y}(x|y) = \frac{p_Z(x,y)}{p_Y(y)} = \binom{k}{x} y^x (1-y)^{k-x} = \text{Binom}(x|k,y). \tag{16}$$

We can describe our computation as a model:

- $Y$ = coin's probability of heads
- $X$ = observed number of heads out of $k$ independent coin flips
- $Y = y \sim \text{Uniform}(y|0,1)$
- $X = x | Y = y \sim \text{Binom}(x|k,y)$.

Suppose we modify the distribution of $Y$ (imagine we have another coin factory):

$$Y = y \sim \text{Beta}(y|\alpha,\beta) \tag{17}$$

Under the new model, we have

$$p_Z(x,y) = p_{X|Y}(x|y) p_Y(y) = \binom{k}{x} y^x (1-y)^{k-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1}. \tag{18}$$

We want to calculate $p_{Y|X}(y|x)$:

$$p_{Y|X}(y|x) = \frac{p_Z(x,y)}{p_X(x)} \propto y^{x+\alpha-1} (1-y)^{k-x+\beta-1} \tag{19}$$

Note that $y^{x+\alpha-1}(1-y)^{k-x+\beta-1}$ is the kernel of the beta distribution[4]. Hence,

$$p_{Y|X}(y|x) = \frac{\Gamma(k+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(k-x+\beta)} y^{x+\alpha-1} (1-y)^{k-x+\beta-1}. \tag{20}$$

[4] Recognizing the kernel of a distribution can often save some tedious computations.