**12. Joint densities and independence** (Friday Feb 5) LM 3.7

**12.1 Joint probability mass functions** (LM P.203-5)

If $X$ and $Y$ are discrete random variables the *joint pmf* is $f_{X,Y}(x,y) \;=\; P(X=x, y=y) \quad \text{for } x \in \mathcal{X},\ y \in \mathcal{Y}$.
Then the *marginal pmfs* of $X$ and of $Y$ are

$$f_X(x) \;=\; P(X=x) \;=\; \sum_{y\in\mathcal{Y}} f_{X,Y}(x,y), \quad \text{and } f_Y(y) \;=\; \sum_{x\in\mathcal{X}} f_{X,Y}(x,y).$$

Note $f_X(x) > 0$ for $x \in \mathcal{X}$, and $f_Y(y) > 0$ for $y \in \mathcal{Y}$, but $f_{X,Y}(x,y)$ can be 0 for some $x \in \mathcal{X},\ y \in \mathcal{Y}$.

**12.2 Independence of two discrete random variables**

$X$ and $Y$ are *independent* if for any subsets $A$ and $B$ of $\Re$, $P(X \in A \,\cap\, Y \in B) \;=\; P(X \in A) \times P(Y \in B)$.
This is equivalent to $f_{X,Y}(x,y) \;=\; P(X=x, Y=y) \;=\; P(X=x).P(Y=y) \;=\; f_X(x)f_Y(y)$.
Clearly this is *necessary*: take $A = \{x\}$ and $B = \{y\}$.
Conversely, if $f_{X,Y}(x,y) \;=\; f_X(x)f_Y(y)$, then for any $A$, $B$:

$$\begin{aligned} P(X \in A,\ Y \in B) \;&=\; \sum_{x\in A}\sum_{y\in B} f_{X,Y}(x,y) \;=\; \sum_{x\in A}\sum_{y\in B} f_X(x)f_Y(y) \\ &=\; \sum_{x\in A} f_X(x) \sum_{y\in B} f_Y(y) \;=\; P(X\in A)\ P(Y \in B) \end{aligned}$$

Note this must hold for all $x, y$. Thus the ranges of the r.v.s cannot depend on each other.

**12.3 Joint and marginal (cumulative) distribution functions**

For two random variables $X$ and $Y$ the *joint cdf* is $F_{X,Y}(a,b) \;=\; P(X \le a,\ Y \le b),\ \ -\infty < a, b < \infty$.
Note that the *marginal cdfs* of $X$ and of $Y$ are given by

$$\begin{aligned} F_X(a) \;&=\; P(X \le a) \;=\; P(X \le a,\ Y < \infty) \;=\; P(\lim_{b\to\infty}\{w; X \le a, Y \le b\}) \\ &=\; \lim_{b\to\infty} P(X \le a,\ Y \le b) \;=\; \lim_{b\to\infty} F_{X,Y}(a,b) \;\equiv\; F_{X,Y}(a,\infty) \\ F_Y(b) \;&=\; P(Y \le b) \;=\; \lim_{a\to\infty} F_{X,Y}(a,b) \;\equiv\; F_{X,Y}(\infty,b) \end{aligned}$$

Just as in 1 dimension, we can get all other probabilities from $F_{X,Y}$. For example :

$$P(a_1 < X \le a_2,\ b_1 < Y \le b_2) \;=\; F_{X,Y}(a_2,b_2) - F_{X,Y}(a_1,b_2) \;-\; F_{X,Y}(a_2,b_1) \;+\; F_{X,Y}(a_1,b_1)$$

**12.4 Joint and marginal probability density functions**

(i) Random variables $X$ and $Y$ are *jointly continuous* if there is a function $f_{X,Y}(x,y)$ defined for all real $x$ and $y$, such that for every (?) set C in $\Re^2$, $P((X,Y) \in C) \;=\; \int\int_{(x,y)\in C} f_{X,Y}(x,y)\ dx\ dy$.
Then $f_{X,Y}(x,y)$ is the *joint pdf* of $X$ and $Y$.

(ii)

$$F_{X,Y}(a,b) \;=\; P(X \in (-\infty,a],\ Y \in (-\infty,b]) \;=\; \int_{y=-\infty}^{b}\int_{x=-\infty}^{a} f_{X,Y}(x,y)\ dx\ dy$$

$$\text{so } f_{X,Y}(a,b) \;=\; \frac{\partial^2}{\partial a\ \partial b} F_{X,Y}(a,b)$$

(iii)

$$P(X \in A) \;=\; P(X \in A,\ Y \in (-\infty,\infty]) \;=\; \int_{X\in A}\int_{y=-\infty}^{\infty} f_{X,Y}(x,y)\ dx\ dy \;=\; \int_{X\in A} f_X(x)\ dx$$

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

So $f_X(x)$ is marginal pdf of $X$ and similarly marginal pdf of $Y$ is $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$

**13. Likelihood: the pdf of an n-sample** (LM 3.7, 5.2)

**13.1 Independence of two r.vs: continuous case** (LM 3.7)

With $A = (-\infty, x)$ and $B = (-\infty, y)$ we see $F_{X,Y}(x, y) = F_X(x)F_Y(y)$.

As with the 1-dimensional case, this is also sufficient:

$X$ and $Y$ are independent if and only if $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ for all $x, y$.

Differentiating, we see this means $f_{X,Y}(x, y) = f_X(x)f_Y(y)$, and conversely integrating we see these are equivalent. Note again it must hold for all $x$, $y$: the ranges of the r.vs cannot depend on each other.

$X$ and $Y$ are independent if and only if $f_{X,Y} = f_X(x)f_Y(y)$ for all $x, y$.

Also, if $f_{X,Y}(x, y) = g_1(x)g_2(y)$ for all $x, y$, then $X$ and $Y$ are independent, and $f_X(x) \propto g_1(x)$, $f_Y(y) \propto g_2(y)$.

**13.2 The pdf (discrete or continuous) of an n-sample**

All the above extends from 2 random variables to any number.

Suppose we have an $n$-sample from the density (discrete or continuous) $f(x; \theta)$.

That is, $X_1, ..., X_n$ are i.i.d and each $X_i$ has pdf $f(x; \theta)$.

Then the joint density of $(X_1, ..., X_n)$ is

$$f_{(X_1,...,X_n)}(x_1, x_2, ...., x_n) \quad = \quad \prod_{i=1}^{n} f(x_i \; ; \; \theta)$$

For example: $X_1, ..., X_n \sim \mathcal{E}(\lambda)$; $f_{(X_1,...,X_n)}(x_1, x_2, ...., x_n) \;=\; \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \;=\; \lambda^n \exp(-\lambda \sum_{i=1}^{n} \; x_i)$.

**13.3 The likelihood** (LM 5.2)

Recall we have data random variables which we model as having some probabilities which depend on $\theta$.

Recall, our goal is to make inferences about $\theta$.

The pdf of the data outcomes is the only "connection" between our data and $\theta$.

The idea of *likelihood* is to use the probabilities of data under model directly.

**Definition:** If we have data $y$ which is the outcome of data random variable(s) $Y$, then the *likelihood function* is $L(\theta) \;=\; f_Y(y \; ; \; \theta)$

The *likelihood function* is simply the probability (density) of the data, considered as a function of $\theta$. The idea is that values of $\theta$ which give high probability to the data values we observe are more likely – they provide better explanations of the data.

Note the pdf considers each $\theta$, and the relative probabilities of different data outcomes $y$.

The likliehood function compares different values of $\theta$ as explanations of specific observed data $y$.

**13.4 The likelihood and log-likelihood based on an n-sample** (LM 5.2)

Suppose $Y_1, ..., Y_n$ are the data random variables for an $n$-sample from $f(y; \theta)$, and that the ourcome of $Y_i$ is $y_i$ $(i = 1, ..., n)$. Then

$$L_n(\theta) \quad = \quad f_{Y_1,...,Y_n}(y_1, ..., y_n; \theta) \quad = \quad \prod_{i=1}^{n} f(y_i \; ; \; \theta)$$

Products are messy: sums are neater. So instead of the likelihood we often consider the log-likelihood

$$\ell_n(\theta) \quad = \quad \log_e L_n(\theta) \quad = \quad \sum_{i=1}^{n} \log_e f(y_i; \theta)$$

For example: $\mathcal{E}(\lambda)$: $L_n(\lambda) = \lambda^n \exp(-\lambda \sum_i x_i)$ and $\ell_n(\lambda) = n \log(\lambda) - \lambda \sum_i x_i$.